CAPÍTULO 4

Enquadramento Metodológico e Métodos

4.1. ENQUADRAMENTO METODOLÓGICO

Se a metodologia for tomada não como "conjunto de métodos", mas como descrição das "regras gerais da ciência que guiam e condicionam a elaboração dos métodos e modelos de medida", sendo os métodos "as técnicas específicas de concepção, condução e avaliação da pesquisa empírica" (Overton, 2002, p.29), o enquadramento metodológico de uma investigação deve fundamentar as opções quanto aos métodos, técnicas e procedimentos aplicados; importa por isso esclarecer não só as opções metodológicas mas também, e mais fundamentalmente, as opções metametodológicas que as inspiram. Qualquer destas opções (metodológicas e metametodológicas) deriva, contudo, da natureza do problema sob escrutínio.

Assim, neste capítulo parte-se da delimitação do problema para a contextualização da metodologia, seguindo-se a identificação dos métodos, de amostragem (dimensões, funções e indivíduos), de planificação da experiência e de tratamento de dados. No CAPÍTULO 5 procede-se à caracterização detalhada dos procedimentos adoptados em cada uma das fases de implementação da presente investigação.

4.1.1. Problema

Dois paradigmas de investigação e avaliação da inteligência humana são tomados neste trabalho como objecto, o paradigma diferencial e o paradigma sistémico: o paradigma diferencial, que representa uma linha de investigação clássica, a que deu origem às mais divulgadas técnicas de avaliação da inteligência, os "testes de inteligência", e que alcançou a maior projecção e divulgação ao longo do século XX; e o paradigma sistémico, que emerge da postura crítica de diferentes autores em relação ao paradigma diferencial e procura a ampliação do conceito de inteligência, o reconhecimento do seu carácter polimorfo, a compreensão da sua natureza enquanto sistema complexo e a sua investigação num quadro multidisciplinar. Da reflexão epistemológica e teórica dos capítulos precedentes emerge desde logo a consciência de que não estamos perante duas formas contraditórias, antagónicas, dicotómicas de conceber e investigar a inteligência humana: e isto não só porque, pelo seu carácter multifacetado e complexo, qualquer conceito sistémico de inteligência, por definição, subsume (ainda que ultrapassando-a) a noção mais tradicional ou convencional de inteligência geral, como também porque, mesmo nalgumas perspectivas mais clássicas de inteligência (como em Wechsler) se pode reconhecer os contornos de uma concepção sistémica do funcionamento cognitivo. A evolução de perspectivas teóricas e a concomitante evolução metateórica e de visões do mundo, oportunamente assinaladas a propósito dos dois autores que representam neste trabalho as posições diferencial e sistémica, respectivamente Wechsler e Sternberg, são bem ilustrativas das comunalidades entre estas perspectivas, nos planos teórico e epistemológico. Importa levar mais longe esta reflexão e averiguar as comunalidades também no plano metodológico.

Como se viu, Wechsler perfilhou desde cedo uma posição metodológica declaradamente diferencial, partindo da evidência das diferenças individuais na resolução de problemas de natureza diversificada e assumindo que essas diferenças mais não representam do que a manifestação de diferenças inter-individuais numa capacidade global, a inteligência. Enquanto autor das mais divulgadas técnicas diferenciais de medida da inteligência, entre as quais a Escala de Inteligência de Wechsler para Adultos (*WAIS*) (Wechsler, 1955, 1981, 1997a,b), Wechsler adoptou, quer na construção quer na validação das suas técnicas, metodologias oriundas da psicologia diferencial, como sejam, por exemplo, a opção por um índice de medida que reporta o desempenho de um indivíduo ao desempenho típico da população a que pertence, ou a utilização de técnicas de validação como a correlação com outros testes de inteligência ou com critérios externos e, mais tarde, a análise factorial (Wechsler, 1944,1955; ver também Matarazzo, 1972; Marques,1963). Já a caminhada metodológica de Sternberg foi bastante distinta: durante cerca de duas décadas investiu em técnicas laboratoriais mais ou menos sofisticadas, com recurso a taquistoscópios e computadores, tendo em vista partir do estudo da variabilidade inter-estímulos para chegar à identificação de "leis universais", as componentes de

processamento da informação (a análise componencial, método criado por Sternberg, é a melhor ilustração desta óptica de investigação). Só mais tarde, a partir do início da década de 90, vem a adoptar métodos diferenciais, muito designadamente na construção e validação de uma técnica diferencial de avaliação da inteligência fundamentada na Teoria Triárquica, o *Sternberg Triarchic Abilities Test* (*STAT*) (Sternberg, 1993a,b, 2002d,e). Contudo, em resposta ao repto de Cronbach (1957,1975a) às duas disciplinas da psicologia científica (psicologia experimental e psicologia correlacional), Sternberg propõe-se implementar também estudos mais complexos, em que simultaneamente procura analisar fontes de variabilidade inter-tratamentos e inter-indivíduos, um tipo de método designado por Cronbach de *ATI* (*Aptitude X Treatment Interaction*) e aplicado por Sternberg e colaboradores (1999) em contexto educativo. Acresce que o conceito de validade que transparece nos trabalhos de Sternberg, como não poderia deixar de acontecer posto serem bem mais recentes do que os de Wechsler, remete para a noção de complementaridade, melhor, de reciprocidade entre teoria e operacionalização, uma noção introduzida por Messick (1975, 1980) e consagrada na última edição dos *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999; ver também Afonso, 1987a,b; Dickes, Tournois, Flieller & Kop, 1994). A validade encarada não como "propriedade psicométrica" das medidas (para já não dizer dos instrumentos de medida, uma noção há muito ultrapassada), mas antes como processo contínuo, nunca concluído (processo de "validação"), de recolha de evidências empíricas que fundamentem as inferências ou interpretações (a teoria) a partir das medidas, e que fundamentem as medidas a partir da teoria, assenta numa visão contextualista e dialéctica da construção do conhecimento que, embora se possa admitir que pudesse ter sido bem acolhida por Wechsler, a julgar pela tónica contextualista dos seus trabalhos, não se pode afirmar ter caracterizado a sua obra. Apesar dos percursos metodológicos distintos, há que reconhecer sobreposição parcial entre Wechsler e Sternberg, não só nos planos metateórico e teórico, como assinalado no CAPÍTULO 3, mas também no plano metodológico; apenas esta sobreposição parcial legitimaria, aliás, a opção por articular as perspectivas dos dois autores na presente investigação.

Primeiro, ambas as perspectivas da inteligência admitem diferenças individuais quanto aos construtos que postulam. É verdade que enquanto essas diferenças em Wechsler e outros autores clássicos constituem a matriz de partida para a identificação ou confirmação de um factor de inteligência geral, em Sternberg e outros autores sistémicos constituem mera verificação empírica acessória, não fundamental, na construção dos seus modelos. Ainda assim, o reconhecimento da existência de diferenças relevantes do ponto de vista da predição, isto é, passíveis de quantificação, de interpretação e/ou de utilização em contextos de intervenção psicológica, é comum às perspectivas diferencial e sistémica, muito em particular a Wechsler e Sternberg (uma afirmação que não seria aplicável a outros autores de modelos sistémicos como, por exemplo, Gardner).

Decorre, como segunda comunalidade metodológica entre os dois autores, a operacionalização das concepções teóricas em instrumentos de avaliação psicológica: em Wechsler a grandiosa "empresa" de construção e progressivo aperfeiçoamento de técnicas diferenciais para a medida da inteligência, desde a idade pré-escolar até à adultícia avançada, a qual perdurou para lá da morte do próprio autor até à actualidade; em Sternberg a criação de uma nova técnica diferencial fundamentada na Teoria Triárquica, inicialmente organizada em níveis, desde o ensino infantil ao ensino superior e adultos em geral, a qual abriu caminho à concepção e elaboração actual de outras formas de medida da inteligência funcional (baseadas na avaliação do desempenho em situação ou na avaliação dinâmica[1]).

Decorre também, como terceira sobreposição metodológica entre as duas linhas de investigação, a adopção por ambas de métodos diferenciais de estudo metrológico das medidas proporcionadas pelos respectivos testes, entre os quais métodos de análise correlacional, muito designadamente, métodos de análise multivaridada – análise factorial exploratória e confirmatória.

É precisamente no quadro desta óptica de investigação que emerge o problema que se tomou por objecto no presente trabalho e que de seguida se procura delimitar:

1. Uma parte substancial dos esforços de validação da Teoria Triáquica consistiu no estudo da estrutura interna das medidas obtidas com o *Sternberg Triarchic Abilities Test - STAT*, um tipo de estudo que Sternberg designa de "validade interna", por contraste com os estudos de "validade externa"[2] desenvolvidos no domínio da educação (Sternberg, 1997b, 1999f, 2002c, 2003b, 2006b; Sternberg & Grigorenko, 2006; Sternberg, Grigorenko, Ferrari & Clinkenbeard, 1999; Sternberg, Torff & Grigorenko, 1998a,b) [3].
2. Os mais importantes estudos de "validade interna" (ou de "validação intra-conceito") (Sternberg, Castejón, Prieto, Hautamäki, & Grigorenko, 2001; Sternberg, Ferrari, Clinkenbeard, & Grigorenko, 1996; Sternberg, Grigorenko, Ferrari, & Clinkenbeard, 1999) envolvem análises factoriais exploratórias e confirmatórias,

[1] Este assunto será tratado em maior detalhe adiante (CAPÍTULO 9).

[2] As noções de "validade interna" e de "validade externa", utilizadas repetidamente por Sternberg, não sendo oficialmente estabelecidas e definidas, pelo menos até à última edição dos *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999), parecem corresponder às noções de "validação intra-conceito" e de "validação inter-conceito", respectivamente, introduzidas na obra *La psychometrie*, publicada em 1994 por Paul Dickes, Jocelyne Tournois, André Flieller e Jean-Luc Kop. É esse o sentido que lhes será atribuído no presente texto.

[3] Nos estudos de "validade externa", a avaliação das aptidões, a instrução e a avaliação dos conhecimentos contemplam, cada uma, as três áreas de processamento mental identificadas pela teoria – inteligências analítica, prática e criativa; os estudos envolvem uma de duas estratégias de inserção curricular, adição ou infusão; e os programas de instrução, já implementados e em curso, abrangem uma gama variada de destinatários, desde o ensino básico ao universitário, e de conteúdos, desde a aprendizagem da leitura ao ensino da psicologia.

estas últimas testando o ajustamento dos dados obtidos com o *STAT* a vários modelos derivados de propostas teóricas no domínio da inteligência, atrás caracterizadas – teoria do factor geral, teoria das aptidões múltiplas (verbal, numérica, espacial), teoria triárquica (inteligências analítica, criativa e prática) ou modelos hierárquicos.

3. Um dos modelos testados nos estudos de análise factorial confirmatória decorre directamente da posição crítica de Sternberg à medida tradicional da inteligência: trata-se do modelo do factor geral (*g*). Recorde-se que, de acordo com Sternberg (e à semelhança dos outros autores sistémicos do domínio da inteligência), *g* apenas emerge na análise factorial dos testes tradicionais de inteligência porque eles medem uma gama estreita do funcionamento cognitivo, isto é, apenas uma parte da inteligência – a inteligência analítica – deixando por avaliar outros aspectos do construto, como as inteligências criativa e prática. Ao ampliar a gama de funcionamento avaliada pelos testes, o autor espera que *g* deixe de emergir como factor comum, correspondendo tão só a um dos três factores que, à luz da Teoria Triárquica, espera encontrar: o factor de inteligência analítica.
4. Embora Sternberg e colaboradores, nos supracitados estudos, relatem resultados que consideram coerentes com as expectativas delineadas a partir da Teoria Triárquica[4], os seus trabalhos foram alvo de acesa contestação, tendo mesmo merecido a atenção de um número especial da conceituada revista *Intelligence* [(2003)*, 31*(4)]. Nathan Brody, num dos artigos publicados nesse número (Brody, 2003a), conclui o seguinte:
    - "A análise do ajustamento aos modelos não exclui a possibilidade da presença de variância de *g* no *STAT*. […] A análise confirmatória utilizada por Sternberg e colaboradores não providencia o método ideal para confirmar a presença de variância de *g* no teste. Uma análise do *STAT* juntamente com diversas outras medidas de inteligência forneceria informação adicional sobre o *locus* das aptidões medidas pelo *STAT* na estrutura taxonómica da inteligência. Sternberg e colaboradores (2001) afirmam que a ubiquidade da evidência de *g* na investigação taxonómica compreensiva das aptidões efectuada por Carroll (1993) é atribuível ao seu fracasso relativamente à inclusão das aptidões criativa e prática.

[4] Os resultados destes estudos serão abordados em detalhe a propósito da discussão dos resultados da presente investigação (CAPÍTULO 8)

Idealmente, esta hipótese deveria ser testada por análises confirmatórias de medidas convencionais, que reconhecidamente confirmam o *factor g*, em conjunto com as medidas do *STAT*. Se Sternberg estiver correcto, a aptidão analítica deverá ter, na taxonomia das aptidões, um *locus* diferente das aptidões criativa e prática. A primeira deverá ser altamente saturada em *g*, ao contrário das outras duas."

5. A correlação entre o *STAT* e outros testes de inteligência foi efectuada por Sternberg e colaboradores (1996) mas apenas tomando quatro testes, "do mais convencional para o menos convencional", o *Concept Mastery Test*, o *Watson-Glaser Critical Thinking Appraisal*, o *Cattell Culture-Fair Test of g* e os *Creative Insight Problems* – tendo-se concluído que "quanto menos convencional é um teste, mais elevada é a sua correlação com os subtestes criativos, por comparação com a correlação com os testes analíticos. A correlação com os menos convencionais dos nossos próprios testes, os práticos, foi sempre a mais baixa." (Sternberg et al., 1996, p.132). Assinale-se que neste tipo de estudos não foi ensaiada a correlação entre o *STAT* e qualquer das grandes baterias de avaliação da inteligência, instrumentos que pela heterogeneidade de conteúdo e pela tentativa de representatividade em relação aos problemas, tarefas e situações do quotidiano poderão de algum modo abranger um mais amplo espectro do funcionamento cognitivo, ainda que sendo consensualmente aceites como bons instrumentos de avaliação da inteligência geral ou *g* (Sternberg, 1996, p.132).
6. Apesar da grande diversidade de conteúdos e de tarefas envolvidos nos testes de Wechsler, a análise factorial confirmatória, aplicada como técnica de validação de construto nas últimas edições (Wechsler, 1991, 1992, 1997b, 1999b, 2003), identifica sistematicamente um factor geral, comum a todos os subtestes das baterias, e uma estrutura hierárquica em que, num segundo nível, emergem factores de grupo, nomeadamente os factores de compreensão verbal, organização perceptiva, memória de trabalho e/ou velocidade de processamento. A emergência de *g* é, no manual destes testes, interpretada como evidência de validade de construto, uma vez que, como se viu (CAPÍTULO 3), as baterias de Wechsler pretendem a avaliação da inteligência enquanto construto global. Estes testes oferecem, por isso, a vantagem de constituírem medidas da inteligência suficientemente saturadas em *g* para representarem a concepção tradicional ou

convencional de medida da inteligência[5], mas suficientemente hierarquizadas factorialmente para remeterem para uma estrutura das aptidões equivalente à sugerida por Carroll, no que propiciam a eventual localização das inteligências medidas pelo *STAT* numa taxonomia das aptidões humanas; além disso, por constituírem técnicas de avaliação da inteligência de conteúdo heterogéneo sugerem o estudo das relações entre as inteligências propostas por Sternberg e as diversas formas de medida da inteligência adoptadas por Wechsler e consubstanciadas nos subtestes das suas baterias, uma possibilidade de análise vedada quando se tomem para estudo testes de conteúdo homogéneo, como os chamados "testes de *factor g*" (testes do tipo dos utilizados por Sternberg e colaboradores).

Em face do precedente, afigurou-se interessante o estudo das relações entre medidas compósitas da inteligência global e da inteligência funcional, quer do ponto de vista da validação das técnicas que operacionalizam as concepções de Wechsler e de Sternberg – a *WAIS* e o *STAT* –, quer do ponto de vista da validação das próprias concepções de inteligência. Numa perspectiva ampla e dialéctica de validação, inspirada em Messick (1975, 1980) e fundamentada na última edição dos *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999), o estudo da relação entre medidas proporcionadas por estes dois instrumentos de avaliação da inteligência abre, assim, a possibilidade de aproximação entre dois métodos diferenciais, entre as respectivas concepções de inteligência, e mesmo, entre dois paradigmas de investigação e avaliação da inteligência, cujo paralelismo metateórico se explorou na primeira parte deste trabalho.

4.1.2. Natureza e Enquadramento Metodológico do Problema

Delimitado o problema sob estudo, importa analisar a sua natureza e enquadramento metodológico e metametodológico. Para o efeito, mostra-se útil tomar como referência o modelo metodológico geral proposto por W. Stern, tão cedo quanto 1911, na obra de título *La Psychologie différentielle dans ses fondements méthodologiques* (citada em Gilles, 1999, p.59). A TABELA 4.1 representa, de forma adaptada, a tabela "indivíduos X características ou variáveis" que serviu de ponto de partida a Stern para a identificação de quatro perspectivas metodológicas de investigação diferencial, às quais se acrescentam agora duas, num propósito de actualização. O interesse desta classificação de Stern consiste em fazer notar que a partir dos mesmos dados, das mesmas

[5] Estas baterias são, regra geral, consideradas como os mais "típicos" testes de inteligência, entre outras razões porque fornecem um resultado compósito na forma de QI, razão por que são com frequência e, acrescente-se, de forma bastante simplista, designadas como "testes de QI".

TABELA 4.1

Tabela Indivíduos X Variáveis (exemplo)
(Adaptado de Stern, 1911 citado em Gilles, 1999)

	VARIÁVEIS						
INDIVÍDUOS	**Idade**	**Sexo**	**Escolaridade** (1)	**QI**(2) (*WAIS-III*)	**Inteligência Analítica** (RB) (3)	**Inteligência criativa** (RB) (3)	**(…)**
1	20	1	6	112	12	12	… → **3**
2	19	1	5	118	14	11	…
3	17	1	4	106	10	13	…
4	46	1	5	105	10	13	…
5	17	2	4	72	6	8	… → **4**
6	20	1	5	124	11	11	…
7	19	1	5	101	13	13	…
8	20	2	4	101	9	12	…
(…)	…	…	…	…	…	…	…
				1	**2**		**5 / 6**

PERSPECTIVAS METODOLÓGICAS DE INVESTIGAÇÃO DIFERENCIAL	INDIVÍDUOS	VARIÁVEIS	MOMENTOS (4)	MÉTODO
1	Diversos	Uma	Um	**Estudos de Variação** (Inter-individuais)
	Diversos (agrupados)	Uma	Um	(Inter-grupais) (ex: diferenças entre sexos)
2	Diversos	Diversas	Um	**Estudos Correlacionais** (correlações entre variáveis)
3	Um	Diversas	Um	**Estudos de Perfil ("Psicografia")** (Intra-individuais)
4	Diversos	Diversas	Um	**Estudos Tipológicos** (correlações entre indivíduos)
5	Um	Uma / Diversas	Diversos	**Estudos de Caso** (Intra-individuais)
6	Diversos	Uma / Diversas	Diversos	**Estudos Longitudinais**

(1) 4.: 9 a 11 anos de escolaridade; 5: 12 a 14 anos de escolaridade; 6: 15 a 16 anos de escolaridade.
(2) Resultados padronizados (média 100 e desvio-padrão 15) a partir dos parâmetros da própria amostra (N=250).
(3) Resultados Brutos numa escala de 0 a 15 pontos.
(4) As células cinzentas resultam de acrescentar a instância temporal ao esquema inicial de Stern, que apenas contemplava indivíduos e variáveis. Pode ainda conceber-se a desmultiplicação do esquema pelo acrescento da instância sitiuacional.

informações, é possível conceptualizar perspectivas metodológicas diversas, umas ocupadas sobretudo das diferenças inter-individuais (perspectivas 1 e 2) e outras das diferenças intra-individuais (perspectivas 3 e 4). Ao acrescentar ao esquema uma dimensão temporal, que compreensivelmente Stern em 1911 não incluiu, outras estratégias de investigação, características da segunda metade do século XX, emergem (as perspectivas 5 e 6) uma sobretudo ocupada das diferenças intra-individuais (5), a outra, das diferenças inter-individuais (6). A não inclusão da dimensão temporal chama a atenção para o carácter estático muitas vezes atribuído à psicologia diferencial clássica – cuja metodologia se definia a partir de dois sistemas interligados, o dos grupos teóricos (indivíduos) e o das dimensões psicológicas (variáveis) (Reuchlin, 1964a), e partia do pressuposto da estabilidade temporal das diferenças individuais. A psicologia do desenvolvimento veio, contudo, por volta de meados do século, trazer a lume a universalidade do fenómeno do desenvolvimento, ao que se acrescentou ainda a contestação do postulado da consistência comportamental em que se apoiava então a "abordagem traço", fundamental em psicologia diferencial[6]. Daí o enquadramento tendencialmente mecanicista da metodologia diferencial clássica (centrada no estudo dos factores internos das diferenças, como as variáveis ou dimensões psicológicas) e o enquadramento organicista, ou mesmo contextualista/selectivista, para que tende a abordagem diferencial actual, bem ilustrada, por exemplo, pela Teoria da Vicariância (de acordo com a qual importa investigar, entre outros aspectos, os factores responsáveis pelo *desenvolvimento* do reportório de processos vicariantes, a par das *situações* em que determinado(s) processo(s) se mostra(m) mais eficaz(es) (Reuchlin, 1999/2002; ver também Gilles, 1999). É neste sentido que, tal como se assinala em nota à TABELA 4.1, ao esquema de Stern se pode acrescentar não só uma dimensão temporal (desenvolvimentista/organicista) mas ainda uma dimensão situacional (contextualista), o que conduz o esquema metodológico de Stern no sentido de uma verdadeira integração dos domínios diferencial (na acepção clássica – variáveis inter-individuais) e experimental (variáveis inter-situacionais), na senda de Cronbach (1957, 1975a) e da sua proposta de construção de uma psicologia inter-domínios.

No quadro metodológico complexo para que tende actualmente a metodologia diferencial, há que reconhecer o carácter "clássico" do método adoptado na presente investigação: trata-se do método correlacional multivariado, que consiste no estudo da relação entre um conjunto de variáveis tendo em vista a identificação das ligações entre elas e, mais precisamente, a maneira como se organizam em estruturas latentes, subjacentes às diferenças individuais. O que está em causa é tão só testar o ajustamento dessas estruturas latentes a modelos conceptuais derivados da teoria psicológica, muito

[6] Alude-se aqui à importante obra publicada em 1968 por W.Mischel, *Personality and assessment* (New YorK: John Willey & Sons) na qual, ao sublinhar a variabilidade inter-situacional do comportamento, põe em causa a abordagem traço, baseada no postulado da consistência comportamental (uma posição que veio a ser designada "situacionismo", por contraste com "personalismo").

em particular de teorias da inteligência representativas dos paradigmas diferencial e sistémico; trata-se de um tipo de estudo de "validação intra-conceito" (Dickes, Tournois, Flieller & Kop, 1994, p.50), situado numa tradição axiomática e orientado por uma "teoria definitória" (neste caso, a Teoria Triárquica) que estipula determinada configuração para a estrutura interna das variáveis estudadas. Note-se que o facto de reconhecer a natureza "clássica" da metodologia adoptada não significa que se lhe atribua menor importância ou pertinência: de facto, dificilmente uma só investigação poderá dar conta das quatro fontes de variação identificadas – indivíduos, variáveis, momentos e situações – e todos os tipos de métodos, ou de perspectivas metodológicas de investigação, se mostram úteis e heurísticos para o esclarecimento do papel dessas fontes de variação no comportamento humano. A adopção de um esquema metodológico compreensivo numa psicologia inter-domínios não obriga a que toda e qualquer investigação psicológica seja compreensiva e dê conta de todas as vertentes do esquema metodológico descrito; pelo contrário, um tal esquema constitui antes uma estrutura de referência que permite reconhecer o lugar e o valor de um determinado tipo de método para um propósito específico, em cada domínio de investigação (daí a designação "psicologia inter-domínios"). Assim, não houve neste trabalho preocupação nem com o estudo da instância temporal (os dados foram recolhidos, para cada indivíduo, numa única ocasião), nem com o estudo da instância situacional (os dados foram recolhidos pela aplicação estandardizada das técnicas diferenciais sob estudo), mesmo que reconhecendo o valor inestimável da investigação em qualquer das duas perspectivas. Foram antes recolhidos dados com duas técnicas de avaliação psicológica (cada uma delas medindo uma diversidade de variáveis) numa amostra de indivíduos da população portuguesa, pelo que o estudo lidou essencialmente com duas fontes de variação ou instâncias do esquema metodológico apresentado, os indivíduos e as variáveis, as mais clássicas, afinal, da abordagem diferencial.

Importa então questionar o estatuto metametodológico desta investigação: não será esta uma perspectiva demasiado estreita para investigação de um fenómeno altamente complexo como a inteligência humana? Não se estará a fragmentar o fenómeno sob estudo, ao ignorar as instâncias temporal e situacional, destituindo-o do seu carácter complexo e dinâmico? A resposta a estas questões apela à reflexão sobre o método correlacional quanto aos níveis de observação e explicação para que contribui (Reuchlin, 1999/2002), quanto à visão (ou visões) do mundo a que se reporta (Pepper, 1942, 1966) e, também, quanto ao quadro metametodológico em que se inscreve (Overton, 2002, 2006a).

A estratégia de investigação correlacional assume muito claramente uma perspectiva estrutural de investigação (Reuchlin, 1999/2002) que, ao negar o reducionismo – quer o inerente à perspectiva analítica (que ignora os determinismos descendentes), quer o inerente à perspectiva holista (que ignora os determinismos ascendentes) –, opta por trabalhar simultaneamente com diferentes níveis de

observação e explicação. Admite a existência de elementos constituintes dos fenómenos sob estudo (no que difere da perspectiva holística), mas não se concentra na sua mera identificação ou segregação (no que difere da perspectiva analítica), antes toma por objecto de estudo as inter-relações desses elementos das quais decorre uma "estrutura" ou um "sistema" que contempla simultaneamente vários níveis de observação e explicação. Os métodos de análise multivariada, como a análise factorial, e em particular os métodos estruturais, como o das equações estruturais lineares (técnicas que adiante se apresentam) constituem meios privilegiados de investigação estrutural. Mas é importante não esquecer que a riqueza e o valor heurístico das estruturas que estes métodos fazem emergir dependem em larga medida da representatividade dos fenómenos observados relativamente ao universo de fenómenos envolvidos numa determinada conduta. Ao lidar com duas fontes de variação – os indivíduos e as variáveis – a investigação factorial clássica não pôde senão chegar a estruturas que, ainda que podendo dar conta de diversos níveis de observação e explicação (desde *g* às aptidões específicas) não deixaram ainda assim de lidar apenas com uma parte dos fenómenos relevantes para a compreensão da inteligência, as variáveis do mundo interno do indivíduo. Acresce que os métodos de análise multivariada assentam numa lógica de aditividade de componentes da variância, em que cada resultado é uma função linear de elementos independentes, i.e., o resultado é a soma de efeitos de componentes independentes (Overton, 2003). Não é de surpreender, por isso, que tais modelos não abranjam aspectos desenvolvimentistas ou contextualistas da inteligência e acabem por ser encarados pelos investigadores oriundos de outros paradigmas como estruturas muito estáticas (que ignoram o efeito do desenvolvimento) e rígidas (que ignoram o efeito das situações). Ao optar-se por este tipo de método, e como assinalado acima, tem-se presente esta limitação; mas tem-se também presente que pela natureza sistémica de uma das teorias sob estudo – a Teoria Triárquica da Inteligência – as variáveis envolvidas supõe-se ultrapassarem as da investigação factorial clássica no domínio da inteligência, estando em jogo, de acordo com Sternberg, um mais amplo espectro do funcionamento cognitivo que inclui variáveis experienciais (inteligência criativa) e contextuais (inteligência prática).

A visão do mundo (Pepper, 1942) subjacente à estratégia de investigação correlacional depende, por sua vez, da interpretação que se faz dos factores identificados: estes tanto podem ser tomados como dimensões latentes que "explicam", numa lógica de causalidade linear, o comportamento – e esta é uma interpretação demasiado frequente (por exemplo, o baixo rendimento escolar é "explicado" pelo baixo resultado no teste de inteligência ou pela baixa aptidão verbal); ou podem ser interpretados como configuração, como subsistema de um sistema mais amplo – que reconhece que os factores identificados dizem respeito apenas a uma parte dos fenómenos pertinentes para a compreensão do comportamento – procurando-se a "compreensão" do funcionamento, mais do que a "explicação" do funcionamento. A primeira perspectiva de interpretação dos factores é

mecanicista e remete para uma causalidade aristotélica mecânica ou eficiente (fundada na eficiência do funcionamento de um mecanismo); a segunda é contextualista e remete para uma causalidade aristotélica formal (fundada na forma ou organização do objecto sob estudo) (Overton, 2002; Pepper, 1942; Vasco, 2001).

De acordo com Overton (2002), os métodos correlacionais são com demasiada frequência aplicados numa lógica de causalidade linear, mesmo quando explicitamente é reconhecido que "correlação" não significa "causa". De facto, desde o início da sua formação em psicologia, os estudantes são instruídos e treinados a pensar numa lógica positivista e cartesiana: aprende-se, por exemplo, que quando, sob condições controladas, a administração de um estímulo (antecedente ou variável independente) conduz, de forma quase invariável, a uma resposta ou comportamento (consequente ou variável dependente), fica demonstrado que o estímulo é "causa suficiente" da resposta; ou aprende-se, nas chamadas experiências de privação, que, se à supressão do estímulo (antecedente) corresponde a supressão da resposta (consequente), fica demonstrado que o estímulo é "causa necessária" da resposta. No quadro desta lógica de investigação, característica da "metodologia científica fragmentada" (*split scientific methodology*) (Overton, 2002, p.31), facilmente se adquire o vício de raciocínio de estabelecer relações lineares entre fenómenos, desde que se liguem de algum modo entre si, pelo que as correlações são muitas vezes lidas na óptica de causalidade linear – é o que se passa quando, como no exemplo acima, as estruturas factoriais são interpretadas no quadro da visão do mundo mecanicista (se a inteligência se correlaciona com o sucesso escolar, então este *deve-se* ao, ou *explica-se* pelo, nível de inteligência).

Em reacção à lógica neopositivista e reducionista em que assenta esta postura de investigação, um outro movimento nasceu, o da "compreensão hermenêutica" (*hermeneutic understanding*) (por oposição a "explicação causal") de acordo com o qual "a interpretação é a metodologia que melhor propicia a compreensão" (Overton, 2002, p.37). Enquanto metodologia das ciências sociais e do comportamento, a compreensão hermenêutica centra-se na pessoa e na sua acção, não nas variáveis externamente manipuláveis e observáveis, e por inerência remete para a teleologia ou causalidade aristotélica final (a centração na pessoa em acção obriga ao reconhecimento da intencionalidade dessa acção) e para a teleonomia ou causalidade aristotélica formal (a centração na pessoa obriga ao reconhecimento de uma organização, de padrões de acção).

Estas duas posturas metodológicas – a das causalidades material e eficiente, inspirada nas ciências naturais, e a das causalidades formal e final, oriunda das ciências humanas – estabeleceram-se na ciência psicológica como "guardiãs contra a ansiedade cartesiana" (Overton, 2002, p.38) dando origem a dicotomias como observação *versus* interpretação, explicação *versus* compreensão, ciência natural *versus* ciência social, dicotomias que sugeriram, a partir da década de 50 do século XX, um

esforço de integração ou síntese – a "metodologia científica relacional" – em que estas antinomias são coordenadas como momentos distintos num processo único de investigação empírica. O princípio nuclear da metodologia relacional é o de que "todos os dados são plenos de teoria" (Hanson, 1958 citado em Overton, 2002, p.40): a ideia de complementaridade ou reciprocidade – de dialéctica – entre interpretação e observação elimina a fractura cartesiana entre as duas metodologias e destrói qualquer possibilidade de reducionismo ou fundacionalismo. A interpretação identifica o que será ou não aceite como observação pertinente e as observações determinam o que será ou não admissível como interpretação. "A interpretação sem observação é vazia; a observação sem interpretação é cega" (Overton, 2002, p. 42).

Este quadro de reflexão metametodológico ajuda a entender o sentido e a oportunidade do precioso contributo de Cronbach em 1957, aquando do famoso discurso que proferiu na qualidade de novo presidente da *American Psychological Association*, no qual afirmava a necessidade de construir uma psicologia inter-domínios pelo desenvolvimento de uma metodologia que articulasse as posturas metodológicas experimental (baseada numa lógica neopositivista) e diferencial (baseada numa hermenêutica da compreensão). E ajuda também a enquadrar um outro contributo valioso, o de Messick (1975, 1980) quando sugere, na definição da validade de construto, a ligação dialéctica entre teoria psicológica (interpretação) e medida (observação). Este mesmo quadro de reflexão metametodológico, em ligação com a classificação resultante da ampliação da tabela metodológica proposta por Stern (ver p.252), permite situar o método correlacional da presente investigação claramente no quadro da "hermenêutica da compreensão", posto que não se pretende averiguar relações causais, mas antes identificar configurações ou estruturas que promovam uma melhor compreensão do significado e da natureza do construto inteligência. Permite também reportar a investigação claramente a uma das disciplinas da psicologia científica identificadas por Cronbach, a "psicologia correlacional". E permite ainda conceptualizar esta investigação no quadro da validação de construto na acepção dialéctica sugerida por Messick, em particular da "validação intra-conceito" (Dickes, Tournois, Flieller & Kop, 1999).

Importa, talvez, acrescentar que, como adiante se sugerirá (CAPÍTULO 10), se afigura fundamental para o futuro da investigação da inteligência humana que em toda e qualquer investigação, mesmo adoptando métodos que assumidamente se situam num dos pólos da dicotomia metodológica esboçada por Overton (e denunciada por Cronbach), seja reconhecido o potencial e o carácter indispensável das outras posturas metodológicas, sem o que se correrá o risco de diluir a complexidade do construto inteligência, deixando escapar a oportunidade de construção de modelos verdadeiramente sistémicos, enquadrados numa metateoria relacional.

4.2. CARACTERIZAÇÃO DOS MÉTODOS

De acordo com o precedente, na presente investigação foi adoptada uma perspectiva de investigação diferencial clássica, a estratégia correlacional, "clássica" no sentido em que retém as mais tradicionais fontes de variação identificadas por Stern (1911, citado em Gilles, 1999) na sua classificação metodológica (cf. p.252), as variáveis e os indivíduos, e recorre principalmente a métodos de análise multivariada típicos da psicologia correlacional ou diferencial. A observação foi, assim, efectuada a partir da aplicação de duas técnicas diferenciais, cada uma delas medindo uma diversidade de variáveis, numa amostra de indivíduos da população portuguesa. Ambas as técnicas diferenciais – a *WAIS-III* (*Wechsler Adult Intelligence Scale – Third Edition*) e o *STAT-R (H)* [*Sternberg Triarchic Abilities Test – Revised (Level H: College & University and Adults*)] – foram traduzidas e adaptadas no âmbito do presente trabalho[7], o que implicou todo um conjunto de estudos experimentais prévios. Assim, para além da amostra do Estudo Principal, outras amostras foram examinadas nos Ensaios Experimentais da *WAIS-III* e do *STAT-R (H)*.

A caracterização do método toma por referência a descrição da metodologia diferencial de acordo com Miranda (1981), e inicia-se pela observação – a delimitação das variáveis (amostragem de dimensões e de funções) e respectivos instrumentos de observação, e das técnicas de selecção dos indivíduos (amostragem de indivíduos) – e prossegue com a organização dos planos experimentais – identificação das hipóteses experimentais e organização global da experiência – terminando com o tratamento dos resultados – identificação das técnicas de análise de dados aplicadas.

4.2.1. Amostragem de Dimensões e de Funções

A opção por instrumentos de observação já existentes simplifica as tarefas, fundamentais em investigação diferencial, de amostragem de dimensões e de funções. Tratando-se de instrumentos enquadrados pelas concepções teóricas de dois autores sobre a inteligência, a amostragem de dimensões remete, evidentemente, para o quadro conceptual de cada um, cada instrumento consistindo numa operacionalização da respectiva concepção de inteligência. Do enorme universo de acepções, definições e concepções de inteligência aflorado no CAPÍTULO 1, a amostragem de dimensões no presente trabalho consistiu em optar por duas teorias ou modelos – concepção de inteligência global de Wechsler e concepção de inteligência funcional de Sternberg –, cada uma delas

[7] A tradução das técnicas diferenciais em causa foi autorizada, pela empresa editora da *WAIS-III* (Copyright © 1997 by The Psychological Corporation, a Harcourt Assessment Company. Direitos de tradução portuguesa, Maria João Afonso © 2002 by The Psychological Corporation, a Harcourt Assessment Company. Adaptação e reprodução autorizadas. Todos os direitos reservados) e pelo próprio Prof. Robert Sternberg, autor do *STAT-R (H),* um teste que não ultrapassou ainda a fase experimental.

representativa de um paradigma de investigação da inteligência humana, respectivamente o diferencial e o sistémico, opção cujo fundamento se procurou estabelecer nos três capítulos da PARTE 1. Por outro lado, porque os dois instrumentos foram criados tendo em vista a operacionalização das concepções teóricas dos dois autores, comportam conteúdos destinados à avaliação de funções que cada um considerou, ou estimou, pertinentes para a avaliação dos seus construtos. Acresce que os estudos de validação publicados (sobre a *WAIS-III* ver Wechsler, 1997b; Taub, McGrew & Witta, 2004; Ward, Ryan & Axelrod, 2000; e sobre o *STAT-R (H)* ver Sternberg, Castejón, Prieto, Hautamäki, & Grigorenko, 2001; Sternberg, Ferrari, Clinkenbeard, & Grigorenko, 1996; Sternberg, Grigorenko, Ferrari, & Clinkenbeard, 1999; Sternberg & The Rainbow Project Colaborators, 2004, 2006; ver Sternberg, 2003h, 2006b) sugerem, em ambos os casos, a adequação das medidas à avaliação das dimensões que se propõem avaliarem, o que significa, por outras palavras, que esses instrumentos, nas suas edições originais, são encarados como possibilitando uma adequada amostragem das funções pertinentes para a avaliação dos respectivos construtos (validação de construto).

No início deste projecto de investigação, não existia versão portuguesa de qualquer dos instrumentos escolhidos, pelo que houve necessidade de começar por os traduzir, adaptar e ensaiar em amostras portuguesas, para averiguação das propriedades metrológicas das medidas, tendo em vista aperfeiçoá-los, dentro do possível, e prepará-los para utilização no Estudo Principal. Nesta fase da investigação, procurou-se orientar as opções metodológicas pelos princípios consagrados nas directrizes internacionais relativas à tradução e adaptação de técnicas diferenciais (*International Test Commission, 2000*; ver também Hambleton, 1994; Van de Vijver & Hambleton, 1996) e aplicar métodos de análise de dados generalizadamente aceites pelos especialistas na construção e estudo metrológico de técnicas diferenciais (Aiken & Groth-Marnat, 2006; Anastasi & Urbina, 1997; Cronbach, 1970; Guilford, 1954, Guilford & Fruchter, 1978; Kerlinger, 1973; Kline, 1993; Nunnally, 1978; Nunnally & Bernstein, 1994). A identificação e descrição das fases prévias de tradução, adaptação e ensaio experimental, para cada um dos instrumentos, serão tratadas no CAPÍTULO 5, juntamente com a descrição dos procedimentos adoptados na implementação do Estudo Principal. Para prosseguir na caracterização e fundamentação dos métodos, torna-se necessário por ora identificar e descrever as variáveis de que se partiu para o estudo do problema atrás enunciado, o que implica a descrição dos instrumentos na sua versão final, aplicada no Estudo Principal desta investigação.

1) ESCALA DE INTELIGÊNCIA DE WECHSLER PARA ADULTOS – 3ª EDIÇÃO (*WAIS-III*)

A *WAIS-III* é, de entre as provas de Wechsler, a mais directa herdeira da primeira escala construída pelo autor, a *Wechsler-Bellevue I* (1939) e corresponde à sua terceira revisão – a primeira revisão foi publicada em 1955 e adoptou pela primeira vez a designação de *WAIS*, e a segunda foi

publicada em 1981, o ano da morte de Wechsler, com a designação *WAIS-R*. A última edição, publicada nos Estados Unidos em 1997 destina-se a ser utilizada entre os 16 e os 89 anos e mantém ainda muitas características da primeira edição da prova (Afonso, 2003e, 2004a,e), as quais, aliás, são partilhadas com as escalas de inteligência concebidas por Wechsler para a avaliação da inteligência noutros níveis etários: é uma bateria compósita de aplicação individual, subdividida em duas partes,

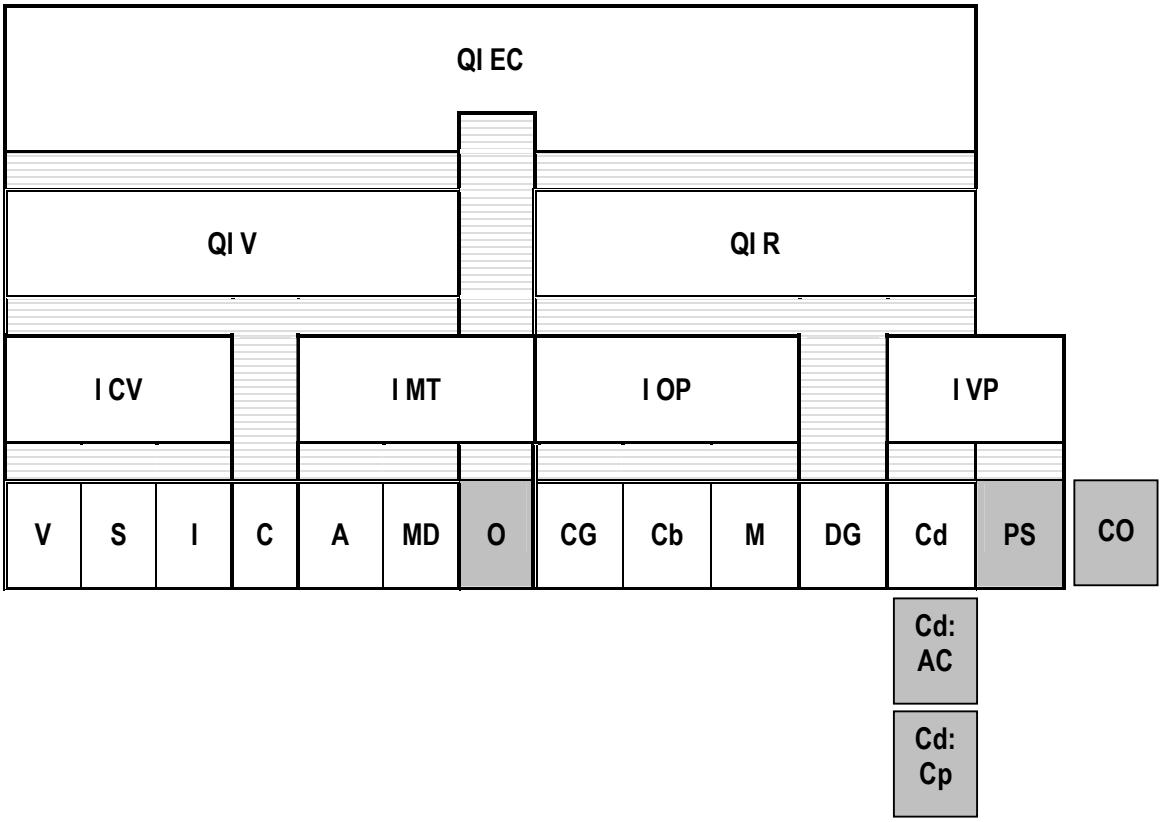

FIGURA 4.1. Estrutura da *WAIS-III*: tipos de resultados

Verbal e Realização, organizada por pontos (não por idades) e que proporciona um resultado global – Quociente de Inteligência (QI) de Escala Completa – e dois resultados parciais – os QIs Verbal e de Realização. Todos estes resultados se expressam em escala intervalar, com média 100 e desvio-padrão 15, e representam, por isso, a distância do resultado individual à média da população, expressa em unidades de desvio-padrão, pelo que se designam habitualmente de QIs de desvio (sobre origem e significado do QI de desvio ver CAPÍTULO 3, p.166). A par dos resultados globais, a escala proporciona ainda resultados para cada subteste, expressos numa distribuição normal com média 10 e desvio-padrão 3, habitualmente designados de resultados normalizados-padronizados (RNP). Na última edição

foram ainda introduzidos Índices Factoriais (IF) – Índices de Compreensão Verbal (ICV), de Organização Perceptiva (IOP), de Memória de Trabalho (IMT) e de Velocidade de Processamento (IVP) – que agrupam conjuntos menores de subtestes e se baseiam na estrutura das medidas identificada por análise factorial; estes índices expressam-se numa escala idêntica à dos QIs (Wechsler, 1997). A FIGURA 4.1 apresenta os vários tipos de resultados e suas relações, ou seja, a estrutura da *WAIS-III*.

TABELA 4.2

Lista dos Subtestes e Resultados Compósitos da *WAIS-III* e respectivas abreviaturas

Subtestes Verbais	Abreviatura	**Subtestes de Realização**	Abreviatura
Vocabulário (2)	**V**	Completamento de Gravuras (1)	**CG**
Semelhanças (4)	**S**	Cubos (5)	**Cb**
Informação (9)	**I**	Matrizes (7)	**M**
Compreensão (11)	**C**	Disposição de Gravuras (10)	**DG**
Aritmética (6)	**A**	Código: Dígito-Símbolo (3)	**Cd**
Memória de Dígitos (8)	**MD**	(Código:Aprendizagem Casual) (3a) [1]	**Cd:AC**
Ordenação de Letras e Números (13)	**O**	(Código Cópia) (15) [1]	**Cd:Cp**
		Pesquisa de Símbolos (12)	**PS**
		(Composição de Objectos) (14)	**CO**
QI Verbal	**QIV**	QI de Realização	**QIR**
Índice de Compreensão Verbal	**ICV**	Índice de Organização Perceptiva	**IOP**
Índice de Memória de Trabalho	**IMT**	Índice de Velocidade de Processamento	**IVP**
QI de Escala Completa	**QI EC**		

[1] Só se aplica em caso de manifesta dificuldade na resolução do subteste de Código: Dígito-Símbolo.

A bateria é composta por catorze subtestes, dos quais, numa aplicação regular, são aplicados onze. Os subtestes que estão representados a cinzento são opcionais e só se aplicam no caso de se querer trabalhar com os resultados do índice factorial, de nível superior, para que contribuem ou de se pretender proceder à averiguação de um desempenho fraco num ou mais subtestes obrigatórios. Assim, a opção relativa aos subtestes a aplicar baseia-se nos resultados que se pretende apurar – Índices Factoriais, QIs ou ambos. Quer se pretenda obter só QIs, ou só Índices Factoriais, há que aplicar obrigatoriamente onze subtestes, embora não precisamente os mesmos (nove são comuns); e para se trabalhar com os dois tipos de resultados, é obrigatória a aplicação de treze subtestes. Os subtestes que na figura se situam no exterior da estrutura dos resultados globais não contribuem para o seu apuramento e destinam-se apenas à averiguação de potencialidades ou défices em áreas específicas de funcionamento. Para simplificação da representação gráfica, na figura os subtestes

são representados pelas respectivas iniciais. Na TABELA 4.2 encontra-se a lista completa dos subtestes e dos resultados compósitos que fazem parte de *WAIS-III*, acompanhados do número de ordem na aplicação[8] e das respectivas abreviaturas (que serão por vezes utilizadas em tabelas e quadros ao longo do texto).

Apesar do muito que partilha com a edição anterior, a *WAIS-III* apresenta importantes inovações (Wechsler, 1997a,b, 2002a,b; ver também Afonso, 2004a,e):

- quanto aos destinatários, foi ampliado o limite etário superior, de 74 anos e 0 meses na *WAIS-R* para 89 anos e 0 meses na *WAIS-III*, o que de certo modo responde à necessidade, progressivamente mais premente em função do aumento da esperança de vida, de instrumentos de avaliação destinados a idades avançadas;
- quanto ao conteúdo, assinala-se o aparecimento de novos subtestes – Ordenação de Letras e Números, Matrizes e Pesquisa de Símbolos – a passagem a opcional do subteste de Composição de Objectos (que anteriormente fazia parte da lista dos subtestes obrigatórios da parte de Realização), a aplicação de dois procedimentos opcionais no subteste de Código: Dígito-Símbolo – Aprendizagem Casual e Cópia – e a inclusão, substituição ou modificação do conteúdo de itens, em particular a extensão de vários subtestes para melhorar a discriminação nos níveis inferiores de aptidão;
- quanto à estrutura, a prova é nesta edição mais flexível, pela possibilidade de optar por aplicar entre onze e catorze subtestes, em função dos resultados pretendidos, da natureza do problema de avaliação ou de variáveis contextuais (como tempo disponível);
- quanto aos tipos de resultados, para além dos resultados normalizados-padronizados dos subtestes e dos tradicionais Quocientes de Inteligência, introduz os Índices Factoriais que abrem novas possibilidades de interpretação, fundamentadas na estrutura interna das medidas;
- quanto à aplicação da prova, salienta-se principalmente a preocupação com a uniformização dos procedimentos de início e de inversão na aplicação dos subtestes: a "inversão" consiste na aplicação, por ordem decrescente e até

[8] Por clareza de exposição, optou-se pela ordenação dos subtestes em função da estrutura factorial da bateria, de acordo com a estrutura da *WAIS-III* apresentada na FIGURA 4.1, a qual não corresponde à ordem de aplicação. Esta é, por isso, indicada entre parêntesis a seguir à designação de cada subteste.

atingidos dois sucessos consecutivos, dos itens iniciais do teste, itens de mais baixo nível de dificuldade, quando haja insucesso nos primeiros itens aplicados;

- quanto ao apuramento e à análise de resultados, adopta-se a conversão dos resultados dos subtestes por comparação com o grupo etário – substituindo o procedimento anterior, em que os resultados dos subtestes eram convertidos por comparação com um grupo de rendimento máximo, com idades compreendidas entre os 20 e os 34 anos; além disso, ampliam-se as propostas de análise e interpretação, por surgirem no manual vários tipos de tabelas que permitem o enquadramento dos resultados individuais, e das diferenças entre resultados, por referência à respectiva frequência de ocorrência na população; a análise de perfis, assim legitimada, viabiliza uma maior ênfase na abordagem intra-individual, por contraste com a tradicional abordagem inter-individual dos resultados (Kaufman & Lichtenberger, 1999).

Uma comparação das três edições da *WAIS* (Afonso, 2004a), possibilitou uma apreciação do sentido genérico da evolução conceptual e metodológica das escalas de inteligência de Wechsler para adultos. A análise da estrutura e extensão dos manuais (Wechsler, 1955, 1981, 1997a,b) revela que embora a organização geral se mantenha, o número de capítulos e a respectiva extensão foi sofrendo alterações sensíveis: houve um claro aumento da extensão total dos manuais, e de cada parte que os compõe, tornando mesmo necessária a separação da informação em dois volumes aquando da publicação da *WAIS-III* (a título de ilustração, a extensão total dos manuais passou de 110 páginas em 1955 para 217+327=544, nos dois manuais de 1997). Assinala-se também significativo aumento da extensão dos capítulos de Considerações Gerais de Aplicação (de 7 para 36 páginas) e de Instruções de Aplicação (de 43 para 117), que traduz não apenas a ampliação da própria prova, como referido atrás, mas também um maior cuidado de especificação das condições de observação a respeitar. Particularmente interessante é a enorme evolução relativa à bibliografia que espelha a própria evolução da investigação no domínio da inteligência, entre os anos 30 e os anos 90: em 1955 não existe lista de referências, e apenas são citadas 5 obras, em notas de rodapé; em 1997 as listas bibliográficas dos dois manuais, de Administração e Cotação e Técnico incluem, respectivamente, 55 e 391 entradas.

Mais importante é a evolução no conteúdo dos manuais: por um lado, a introdução conceptual é progressivamente mais explícita e elaborada; por outro lado, houve notável progresso nos procedimentos de estudo metrológico adoptados, os quais são objecto de descrição pormenorizada no Manual Técnico da última edição. Houve na publicação da *WAIS-III,* por exemplo, uma notória preocupação de cobertura exaustiva do estudo da precisão, sendo apresentados resultados provenientes do uso de diversos métodos de estimação, aplicados não só aos resultados mas também

às diferenças entre os resultados mais representativos. Quanto à validade, é indesmentível a riqueza das evidências empíricas proporcionadas pelo Manual Técnico da 3ª edição e a actualidade das metodologias adoptadas. Particularmente dignos de destaque são os estudos correlacionais com outras medidas da inteligência, os estudos com grupos especiais (clínicos) e a investigação factorial exploratória e confirmatória para análise da congruência entre a estrutura interna das medidas e a organização da escala. A validade, que merece um capítulo extenso do Manual Técnico na *WAIS-III*, é exactamente um dos conteúdos em que se assinala uma maior evolução desde a primeira edição da prova, o que pelo menos em parte se compreende como resultado do aperfeiçoamento do próprio conceito de validade e das metodologias de estimação ao longo das sucessivas edições – entre 1954 e 1985 – da publicação hoje designada *Standards for Educational and Psychological Testing* (AERA, APA & NCME, 1999).

Em síntese, na análise da evolução da *WAIS*, entre a primeira e a terceira edições, identificam-se claras tendências de actualização, embora mais nítidas no plano metodológico do que teórico: a conceptualização subjacente à escala é progressivamente mais clara, bem como fundamentada em revisões de literatura mais amplas que permitem caracterizar melhor a natureza do construto e estabelecer com mais rigor a sua relação com outros construtos; mas a escala mantém a fidelidade ao conceito global de inteligência do autor e não incorporou, nem procurou incorporar, de modo sistemático, os contributos decorrentes da investigação noutras áreas da psicologia. De facto, a tentativa de ampliação da gama de funcionamento cognitivo abrangida, pela adição de novos subtestes e inclusão de novas variáveis com designações oriundas da Psicologia Cognitiva (como "memória de trabalho" ou "velocidade de processamento"), não é explicitamente equacionada no quadro de uma nova concepção de inteligência, nem sequer justificada no âmbito de algum modelo teórico de organização ou de funcionamento cognitivo, o que obriga os autores que se debruçam sobre a interpretação de resultados da *WAIS-III* (por exemplo, Kaufman & Lichtenberger, 1999) a explicitar eles próprios essas ligações conceptuais. Ao contrário, no plano metodológico a evolução deu-se no sentido de um progressivo aperfeiçoamento pela aplicação de novas técnicas, cuidadosamente apresentadas e justificadas, e pelo claro esforço de alargamento e aprofundamento do estudo empírico da escala (Afonso, 2004a). Os resultados obtidos nesse estudo empírico – precisão e validação (Wechsler, 1997b) – justificam, aliás, amplamente a opção por esta escala de inteligência como técnica diferencial representativa da medida da inteligência global, tal como Wechsler a concebia.

E é precisamente por referência à concepção de Wechsler sobre a medida da inteligência que devem ser encarados os subtestes que incluiu nas suas escalas, bem como os subtestes acrescentados na última edição: a concepção é a de que os vários subtestes não medem formas diferentes de inteligência mas constituem diferentes formas de medir a inteligência. Recorde-se que

Wechsler afirma serem relativamente indiferentes as tarefas que se incluem numa avaliação da inteligência, porquanto, de uma maneira ou de outra, ao confrontar e resolver qualquer situação ou problema o indivíduo irá manifestar sempre, em maior ou menor grau, a inteligência global (Wechsler, 1944). Perante os contornos da presente investigação, contudo, a natureza dos subtestes não é indiferente; pelo contrário, é importante reter as suas características distintivas, uma vez que acabam por corresponder a boa parte das variáveis de partida a tratar, as que representam a medida diferencial da inteligência, e que serão relacionadas com outras variáveis que neste trabalho representarão o paradigma sistémico. Segue-se, por isso, a caracterização de cada subteste quanto a conteúdo[9], classificação e cotação, racional ou fundamento para a inclusão na bateria, consistência interna e lugar na estrutura factorial (Grégoire, 2004; Kaufman & Lichtenberger, 1999; Wechsler, 1997b; Zimmerman & Woo-Sam, 1973).

SUBTESTES VERBAIS

SUBTESTE DE VOCABULÁRIO. É constituído por 33 questões, das quais as primeiras três constituem itens de inversão, e a tarefa do examinado consiste em explicar o significado de palavras da linguagem comum, algumas de utilização menos frequente nos níveis mais avançados do teste. A investigação tem demonstrado que palavras relativamente frequentes (acertadas por 60% ou mais dos indivíduos) constituem os itens com mais elevadas saturações em *g* e em factores de compreensão verbal, razão por que os subtestes de vocabulário incluem geralmente uma maioria de palavras comuns e acessíveis (Grégoire, 2004). A classificação distingue dois níveis de resposta correcta, além da incorrecta, e faz-se com recurso a critérios previamente estabelecidos. Este subteste foi introduzido na *Wechsler-Bellevue I* como subteste alternativo, ou de aplicação facultativa, uma vez que Wechsler o considerava demasiado sensível a factores educacionais e culturais; contudo, os estudos empíricos depressa revelaram o seu valor para a derivação de um índice global de inteligência, pois apresentava as correlações mais elevadas com o QI de Escala Completa e a mais alta saturação em *g,* sendo ao mesmo tempo o que obtinha melhores índices de precisão – na *WAIS-III*, o coeficiente de consistência interna médio das subamostras por grupo etário foi de .93[10]. Por outro lado, as diferenças qualitativas nas respostas apresentam valor clínico, por serem reveladoras da natureza dos processos de pensamento e até de aspectos não-intelectivos, como afectos e motivações. Na edição original da *WAIS-III,* apresentou a mais elevada saturação no factor de Compreensão Verbal (.89), na amostra

[9] De acordo com o contrato celebrado com a empresa editora aquando da aquisição dos direitos de tradução e adaptação, os conteúdos específicos (itens) da *WAIS-III,* ou mesmo de respostas dos examinados, não podem ser explicitamente mencionados em nenhum relatório ou documento produzido no âmbito da presente investigação; por esta razão, omitem-se exemplos de itens ou de respostas.

[10] Calculado pela fórmula recomendada por Guilford (1954) e Nunnally (1978) (citados em Wechsler, 1997, p. 48) e a partir da transformação *z* de Fisher. Todos os índices de consistência interna recolhidos em Wechsler, 1997, p. 50, Tabela 3.1.

geral, e de forma consistente em todos os níveis etários (saturações entre .86 e .92)[11], ao lado dos subtestes de Semelhanças, Informação e Compreensão. Trata-se de uma boa medida de inteligência cristalizada ou *gc,* envolvendo sobretudo a capacidade de elaboração conceptual, e tem uma saturação de .83 em *g* [12].

SUBTESTE DE SEMELHANÇAS. Integra um total de 19 itens, os cinco primeiros de inversão, nos quais o examinado deve indicar qual a relação de semelhança entre dois conceitos expressos verbalmente. A classificação distingue dois níveis de qualidade das respostas correctas, excepto nos itens de inversão que apenas se classificam como certos ou errados, e recorre a critérios de classificação pré-estabelecidos. Wechsler introduziu este subteste na sua bateria por acreditar tratar-se de uma boa medida de inteligência geral ao apelar ao pensamento categorial, à capacidade de abstracção e de formação de conceitos. Em geral, as respostas podem ser de três níveis, reveladores de graus distintos de desenvolvimento cognitivo: de nível concreto (comparação de qualidades sensíveis dos objectos/conceitos), de nível funcional (comparação das funções ou utilidade dos objectos/conceitos) ou de nível abstracto (identificação da classe a que pertencem ambos os objectos/conceitos). Na edição original da *WAIS-III*, este subteste apresentou um índice de consistência interna média de .86 e registou a saturação mais elevada, .76, no factor de Compreensão Verbal, juntamente com Vocabulário, Informação e Compreensão, tendo as saturações nesse factor, nos diferentes níveis etários, oscilado entre .68 e .83. A saturação em *g* foi de .79.

SUBTESTE DE INFORMAÇÃO. Composto por 28 itens, os quatro primeiros de inversão, nos quais são colocadas perguntas abrangendo um amplo leque de conhecimentos gerais, por exemplo, de geografia, de história ou de ciência. As respostas, que são concisas e precisas, apenas são classificadas como certas ou erradas e existe para cada item uma lista das respostas aceitáveis que facilita a classificação, pelo que este é geralmente um dos subtestes com mais elevado nível de precisão nas baterias de Wechsler – na *WAIS-III*, o índice de consistência interna médio dos grupos etários foi de .91. A inclusão deste subteste baseou-se na premissa de que os indivíduos mais inteligentes possuem uma gama de interesses mais ampla, maior curiosidade intelectual e tendem a procurar mais estímulos mentais, de onde resulta um leque mais amplo e aprofundado de conhecimentos. Na base da inclusão deste subteste há assim a noção de que a informação, ou "cultura geral", não é adquirida passivamente, não é uma simples marca deixada pelo meio no indivíduo, mas resulta de actividade cognitiva deliberada e consciente (da interacção indivíduo X meio). Esta ideia

[11] Método: análise factorial em eixos principais, rotação oblíqua, correlações entre factores de .63 a .73 (Wechsler, 1997, pp. 104-105, Tabela 4.17, amostra geral, e pp. 108-109, Tabelas 4.18 a 4.21, amostras por níveis etários). Salvo nota em contrário, todos os comentários aos resultados da análise factorial exploratória da versão original da *WAIS-III* remetem para esta referência bibliográfica e para estas tabelas).

[12] As saturações dos subtestes no factor g não são apresentadas no manual da *WAIS-III*; recorre-se por isso aos valores apresentados por Kaufman & Lichtenberger (1999, p.79).

encontra apoio na evidência empírica de que esta é uma das provas de inteligência mais correlacionadas com um dos cinco factores de personalidade do modelo "*Big Five Factors*" (sobre este modelo, ver, por exemplo, Gilles, 1999), o factor de Abertura à Experiência (*Openness*) caracterizado pela curiosidade e pela procura activa de estímulos e de experiências novas (Grégoire, 2004). Constitui uma das melhores medidas de inteligência cristalizada (*gc*) e tem uma saturação em *g* de .79. Na edição original da *WAIS-III* saturou no factor de Compreensão Verbal, juntamente com os subtestes de Vocabulário, Semelhanças e Compreensão, com uma saturação de .81 na amostra total e saturações entre .71 e .83 nas subamostras por nível etário.

SUBTESTE DE COMPREENSÃO. Comporta um total de 18 itens, os três primeiros de inversão, que levantam questões acerca de problemas da vida quotidiana, costumes ou hábitos culturais e situações de carácter social ou interpessoal. Só é aplicado para a obtenção de QIs, não contribuindo para nenhum Índice Factorial Com excepção dos itens de inversão, cotados como certos ou errados, a classificação distingue dois níveis de resposta em função da sua qualidade e baseia-se em critérios previamente definidos, mas exige neste caso considerável exercício de julgamento por parte de quem classifica, o que em parte justificará ser um dos testes com mais baixa precisão da bateria – na *WAIS-III*, índice de consistência interna de .84. Wechsler terá incluído este subteste na sequência do valor empírico que provas deste tipo haviam demonstrado, desde as escalas de Binet; para além disso, este tipo de teste apresentava potencialidades do ponto de vista clínico, ao revelar os processos de pensamento a par de alguns factores conativos e afectivos do comportamento. Para muitos autores, trata-se de uma medida de inteligência social (Kaufman & Lichtenberger, 1999; Zimmerman & Woo-Sam, 1973), mas Wechsler (1944, p.68) recusava este conceito e considerava que se tratava antes de um teste de "senso comum": o sucesso depende da retenção de informação prática mas, mais ainda, da capacidade geral para avaliar e aproveitar a experiência passada (talvez aquilo que Sternberg mais tarde veio a designar de "conhecimento tácito"). Estudos empíricos não confirmam o valor preditivo dos resultados deste subteste para a adaptação social, posto que, por exemplo, são baixas as suas correlações com escalas de competência social ou com a escala de Psicopatia do *Multiphasic Minnesota Personalty Inventory* (*MMPI*). Representa na *WAIS-III* uma medida saturada em factor de Compreensão Verbal, ao lado dos três testes anteriores, com uma saturação de .80, na amostra global, e entre .76 e .84, nas amostras por nível etário. Este subteste é considerado uma medida típica da inteligência cristalizada (*gc*) e a sua saturação em *g* é de .77.

SUBTESTE DE ARITMÉTICA. Compõe-se de 24 itens, quatro de inversão, e consiste em pequenos problemas cuja resolução exige o cálculo mental (não é utilizado papel e lápis) e a aplicação de uma ou mais operações aritméticas. Com excepção dos últimos dois itens, onde em caso de resposta rápida a cotação pode receber uma bonificação e atingir dois pontos, em todo o restante subteste apenas se

distinguem respostas certas e erradas, estando previamente definidas com toda a objectividade as repostas aceitáveis para cada item. Este tipo de prova fazia já parte dos testes de inteligência existentes quando Wechsler construiu as suas primeiras escalas e terá sido a experiência positiva e a correlação elevada que tendiam a ter com o resultado total dos testes em que eram incluídos que terá contribuído para a sua escolha. Embora faça apelo a alguns conhecimentos adquiridos através da escolarização, pelo que requer um nível mínimo de domínio das quatro operações aritméticas, é um erro considerar que avalia os conhecimentos adquiridos em disciplinas de aritmética ou matemática, pois representa uma amostragem muito pouco representativa dos conteúdos de tais disciplinas. Trata-se antes de uma prova que exige um processamento de elevada complexidade – compreensão do enunciado verbal do problema, codificação e conservação dos dados do problema na memória de trabalho, opção pelas operações a aplicar no tratamento, conservação na memória de trabalho de resultados intermédios, nos problemas mais avançados, e emissão da resposta final. O nível de consistência interna médio na *WAIS-III* foi de .88 e a saturação em *g* é da ordem de .75. Embora em muitos estudos factoriais sature em mais do que um factor, simultaneamente em factores de compreensão verbal e de memória de trabalho (Grégoire, 2004), na *WAIS-III* original saturou juntamente com o subteste de Memória de Dígitos no factor de Memória de Trabalho, ainda que com uma saturação não muito elevada (.51), e muito secundariamente no factor de Compreensão Verbal (.22). Nas amostras separadas por nível etário, as saturações no factor de Memória de Trabalho variaram entre .41 e .56.

SUBTESTE DE MEMÓRIA DE DÍGITOS. É constituído por sequências de dígitos que o examinado deve repetir logo após apresentação oral pelo examinador. Compõe-se de duas partes, a primeira que exige a repetição dos dígitos na ordem em que foram apresentados – Ordem Directa – e a segunda que exige que, ao repetir, o examinado inverta a ordem de emissão dos dígitos – Ordem Inversa; na primeira parte as sequências têm entre 2 e 9 dígitos e na segunda têm entre 2 e 8 dígitos. Para cada item há dois ensaios e a pontuação é o número de ensaios correctamente efectuados. Esta é uma prova relativamente fraca como medida de *g* (saturação de .57) mas apresenta em geral elevados valores de consistência interna – na *WAIS-III*, coeficiente de .90. A razão para a inclusão nas escalas de Wechsler reside na sua utilidade sobretudo nos níveis inferiores de eficiência, na discriminação da deficiência mental – em geral, a retenção de uma sequência não superior a 4 dígitos na ordem directa ou 2 na ordem inversa é indicadora de défice acentuado da função cognitiva [na *WAIS-III*, para o conjunto de todos os grupos etários, as médias dos números de dígitos correctamente repetidos foram, para a Ordem Directa e Inversa, respectivamente, 6.43 (desvio padrão 1.36) e 4.70 (desvio padrão 1.43) (Wechsler, 1997b)]. Por ser pouco sensível ao nível de escolaridade e, inversamente, muito sensível à idade, a prova assume também importância no diagnóstico clínico. A saturação mais

elevada deste subteste da *WAIS-III* registou-se no factor de Memória de Trabalho, .71, juntamente com os subtestes de Aritmética e de Ordenação de Letras e Números. Nas subamostras etárias, as saturações no factor de Memória de Trabalho variaram entre .58 e .79, sendo mais elevadas nos grupos mais jovens.

SUBTESTE DE ORDENAÇÃO DE LETRAS E NÚMEROS. Como assinalado atrás, este é um dos subtestes acrescentados à *WAIS* na sua última edição e é um dos testes opcionais, que apenas se aplica quando se pretende apurar o Índice Factorial de Memória de Trabalho (IMT). Comporta um conjunto de 7 itens com três ensaios cada, em que a tarefa consiste em repetir as sequências de letras e de números que foram apresentados desordenadamente pelo examinador, procedendo à ordenação dos números, primeiro, por ordem crescente e das letras, de seguida, por ordem alfabética. A extensão das séries varia entre um número e uma letra, e quatro números e quatro letras e a cotação corresponde ao número de ensaios correctamente respondidos. A inclusão deste subteste teve em vista o reforço da medida da memória de trabalho e da atenção, ou seja, do factor de Memória de Trabalho, que representava um "factor extra" nas edições anteriores da *WAIS* (Wechsler, 1997b, p.14). Trata-se, contudo, de um subteste que em larga medida se sobrepõe ao de Memória de Dígitos, já que cerca de 68% da sua variância pode ser prevista a partir do resultado deste último. A consistência interna é de .82 e a saturação factorial apresenta valores entre .60 e .79 no factor de Memória de Trabalho, nos vários grupos etários, e de .62 na amostra total, ao lado dos subtestes de Aritmética e Memória de Dígitos. A saturação em *g* é moderada, na ordem de .65.

SUBTESTES DE REALIZAÇÃO

SUBTESTE DE COMPLETAMENTO DE GRAVURAS. É o primeiro subteste da bateria a ser aplicado e preenche, por isso, uma função de adaptação à situação de teste. Utiliza como material um Bloco de Estímulos onde se encontram 25 gravuras incompletas representando objectos comuns, as cinco primeiras correspondentes aos itens de inversão; a tarefa consiste em identificar, nomeando ou simplesmente apontando na gravura, após uma exposição máxima de 20 segundos, qual a parte importante que lhe falta; nuns itens está em causa apenas o reconhecimento do objecto em causa e a comparação da imagem apresentada com a imagem memorizada; outros implicam também algum grau de raciocínio. A classificação apenas distingue respostas certas e erradas. Este tipo de testes fazia já parte de outras escalas de inteligência, geralmente sob a designação de "teste de figuras mutiladas", quando Wechsler optou pela inclusão nas suas baterias; embora se mostre muito pouco discriminativo nos níveis superiores de funcionamento, apresenta interesse clínico nos níveis médio e inferior, razão da sua manutenção ao longo das três edições. De facto, as distribuições de resultados neste subteste tendem a ser assimétricas negativas, ou seja, a registar concentração de resultados no topo superior da distribuição, mas exactamente por isso o teste transmite confiança aos examinados, pela facilidade

de execução aliada ao carácter lúdico da tarefa – essa a justificação para que seja o primeiro subteste a ser aplicado. O nível de consistência interna é de .83 e a saturação em *g* relativamente moderada, .64. Na amostra conjunta de aferição da *WAIS-III*, a saturação mais elevada deste subteste, .56, foi registada no factor de Organização Perceptiva, ao lado de Cubos, Matrizes e Disposição de Gravuras,. Nos diversos níveis etários, as saturações no factor de Organização Perceptiva foram em geral pouco elevadas, entre .41 e .56, e na amostra de idade mais avançada, 75-89 anos, o subteste saturou no factor de Velocidade de Processamento (.62), juntamente com Cubos, Disposição de Gravuras, Código e Pesquisa de Símbolos.

SUBTESTE DE CUBOS. Este subteste é composto por 14 itens, dos quais quatro de inversão, e utiliza como material conjuntos de cubos bicolores que devem ser reunidos pelo examinado de modo a formar padrões semelhantes aos construídos pelo examinador (itens de inversão) ou representados graficamente no Bloco de Estímulos. O nível de dificuldade crescente liga-se não só ao tipo de modelo (tri ou bi-dimensional) e ao número de cubos envolvidos (dois, quatro ou nove), como à passagem da escala do modelo de 1/1, nos itens mais fáceis, para 1/2, nos mais difíceis. Nos primeiros seis itens há a possibilidade de cometer um erro e repetir a tarefa num segundo ensaio; os restantes são itens de ensaio único, mas distinguem quatro níveis de resposta correcta em função do tempo de execução. Esta prova é tradicionalmente a melhor medida não verbal de inteligência de toda a bateria e a ela se junta, na *WAIS-III* o subteste de Matrizes: a saturação em *g* é, nesta edição do teste, de .72. A origem do subteste remonta a 1923, ano em que Kohs criou um teste de Cubos, que veio a ser adaptado para diversas baterias, com o qual pretendia avaliar as capacidades de análise e de síntese sem recurso à linguagem, pois considerava corresponderem, afinal, à noção comum de inteligência (Grégoire, 2004). Para além dessas capacidades, a prova envolve ainda a coordenação visual-motora e a capacidade para trabalhar sob pressão de tempo. Wechsler incluiu este subteste na construção das suas baterias não só por o considerar uma boa medida da inteligência enquanto construto global, como ainda pela possibilidade que fornece de observação dos processos de resolução adoptados pelo examinado, observação de inquestionável valor clínico. Com efeito, a investigação cognitiva veio a identificar três estratégias de resolução neste subteste (já antes assinaladas – CAPÍTULO 1, p.63): na analítica, o examinado subdivide a figura nas suas unidades constituintes, localiza as faces dos cubos que correspondem a cada uma dessas unidades e junta-as para formar o todo; na global, procede por ensaio e erro, sempre na tentativa de construir o padrão total; na sintética, subdivide a figura em grupos menores de cubos, estruturas parciais, e estes tornam-se objectivos intermédios que possibilitam o tratamento de menor quantidade de informação de cada vez. A opção por uma determinada estratégia parece ter ligação com outras variáveis como a dependência/independência de campo de Witkin (estratégia global nos dependentes, estratégia sintética nos independentes) ou a idade (a estratégia

global tende a predominar sobre as outras duas nas idades ou mais jovens, ou mais avançadas) (Grégoire, 2004). Este subteste apresenta geralmente um bom nível de consistência interna média, .86 na *WAIS-III*, e satura claramente no factor de Organização Perceptiva, sendo a mais alta saturação (.71) de entre os subtestes que definem esse factor (além de Cubos, Completamento de Gravuras, Matrizes e Disposição de Gravuras). As saturações nos diferentes níveis etários situam-se entre .59 e .73, mas, tal como o subteste de Completamento de Gravuras, no grupo de idade mais avançada, 75-89 anos, o subteste saturou, ainda que moderadamente (.51), no factor de Velocidade de Processamento.

SUBTESTE DE MATRIZES. Introduzido apenas na *WAIS-III*, por ser um tipo de prova tradicionalmente muito eficaz como medida da inteligência fluida, este subteste é constituído por 26 itens, dos quais três de inversão, precedidos de três exemplos não cotados. O seu conteúdo inspira-se muito directamente no teste de Matrizes Progressivas, criado em 1938 por Raven e sucessivamente aperfeiçoado até aos nossos dias (Raven., Raven & Court, 1995), e compõe-se de quatro tipos de itens: completamento de padrões, classificação, raciocínio analógico e raciocínio serial. A resolução não obedece a limites de tempo e as respostas são emitidas através da opção por uma entre cinco alternativas possíveis. Os itens e as alternativas de resposta encontram-se no Bloco de Estímulos, utilizado na maioria dos subtestes de realização, e as respostas classificam-se apenas como certas ou erradas. Por não atender a limites de tempo, esta prova é particularmente adequada à avaliação das pessoas mais velhas, já que estas são geralmente prejudicadas pelas tarefas que recebem bonificação por desempenhos rápidos; em contrapartida, a sua aplicação representa regra geral uma percentagem importante do tempo total de aplicação. De certo modo, este subteste substitui, numa aplicação regular da *WAIS*, o subteste de Composição de Objectos, que passou a opcional não só pelo tempo longo de aplicação, mas pelos fracos níveis de consistência interna e de saturação em *g*. Do ponto de vista da duração da aplicação, a substituição terá sido inconsequente, como referido; mas do ponto de vista metrológico, parece ter sido francamente favorável, já que o coeficiente de consistência interna médio na *WAIS-III* foi de .90 (muito superior ao de Composição de Objectos, .70) e a saturação em g é .72, (também superior à do subteste de Composição de Objectos, .62). Esta é uma prova altamente sensível à idade e nela o declínio após os 45 anos é evidente, algo coerente com o facto de ser uma boa medida da inteligência fluida; a amostra mais idosa, 75-89 anos, atinge os resultados médios mais baixos de toda a bateria neste subteste. A resolução dos itens desta prova parece depender em alguma medida da memória de trabalho: o examinado começa por decompor as figuras da série que de seguida conserva na memória enquanto descobre a regra subjacente à ordem das figuras. Quanto maior o número de elementos da série, mais difícil se torna a tarefa de mantê-los na memória de trabalho. Talvez por isso, à saturação mais alta deste subteste no factor de Organização Perceptiva,

.61, juntamente com Completamento de Gravuras, Cubos e Disposição de Gravuras, segue-se uma saturação baixa mas assinalável de .21 no factor de Memória de Trabalho. Nas amostras por nível etário as saturações em Organização Perceptiva variam entre .42 e .67, e este é o único teste deste factor que mantém a sua natureza factorial no grupo de idade mais avançada, não passando a saturar no factor de Velocidade de Processamento como o Completamento de Gravuras e os Cubos, o que se compreende por não ser executado com tempo limitado.

SUBTESTE DE DISPOSIÇÃO DE GRAVURAS. Só é aplicado quando se pretende apurar QIs, uma vez que não contribui para nenhum Índice Factorial. É constituído por 11 itens, todos de aplicação obrigatória, que requerem a ordenação de um conjunto de gravuras, representando situações de carácter social, de modo a que contem uma história. As gravuras estão impressas em cartões e formam sequências que variam em extensão entre três (item exemplo) e seis cartões. O tempo para cada item é limitado, mas a pontuação não premeia desempenhos rápidos, antes premeia algumas respostas de acordo com critério previamente estabelecido, o que, em cinco itens, distingue dois níveis de qualidade da resposta. Embora descrendo da existência de uma inteligência social como construto independente da inteligência geral, Wechsler acredita ser útil avaliar a maneira como a inteligência se aplica no contexto social, razão por que inclui este subteste nas suas provas. A investigação tendeu a confirmar que este subteste não é de facto uma medida eficaz de "inteligência social", sendo nulas as correlações quer com testes de competência social, quer com a escala de Psicopatia do *MMPI-2*. Parece antes envolver a distinção entre o essencial e o acessório, a capacidade de integração de elementos numa sequência coerente e a capacidade de planeamento, esta última uma das componentes das funções executivas, pelo que este é um subteste particularmente adequado ao despiste de perturbações a esse nível. Alguns autores (Kaufman & Lichtenberger, 1999) admitem ainda que envolva o pensamento divergente ou a criatividade. O potencial deste subteste do ponto de vista do diagnóstico clínico foi determinante para mantê-lo na bateria, tendo mesmo sido sugerida a sua utilização como técnica projectiva, a partir da verbalização das histórias, para o que foi mesmo desenvolvida uma grelha de classificação (Segal et al., citados em Grégoire, 2004). Aliás, a consistência interna na ordem de .74, uma das mais baixas de toda a bateria, e a moderada saturação em *g*, .66, fazem desta uma das provas menos robustas cuja retenção ao longo das várias edições se justifica sobretudo pelo seu valor clínico. À semelhança de Completamento de Gravuras, de Cubos e de Matrizes, também este subteste da *WAIS-III* contou para a definição do factor de Organização Perceptiva, com saturações entre .41 e .56, excepto para o grupo mais idoso, 75-89 anos, em que satura no factor de Velocidade de Processamento (.47), algo que é interpretado como reflexo da existência de limites de tempo (tal como em Completamento de Gravuras e em Cubos). Contudo, o

nível modesto das saturações em qualquer dos factores justifica que este subteste não contribua para o apuramento dos Índices Factoriais.

SUBTESTE DE CÓDIGO: DÍGITO-SÍMBOLO. Consiste em copiar símbolos associados aos nove algarismos ou dígitos, a partir de uma chave que se mantém visível durante a execução do teste, o mais rápida e exactamente possível, até um limite de 120 segundos. Na *WAIS-III* surgem como novidade os procedimentos opcionais do Código que devem ser aplicados quando se verifiquem dificuldades na resolução do teste base, Código: Dígito-Símbolo. Um dos Procedimentos é o **CÓDIGO: APRENDIZAGEM CASUAL**, um subteste opcional que deve ser aplicado imediatamente a seguir ao subteste de Código e que consiste em duas tarefas, a de **Emparelhamento**, fazer corresponder de memória aos nove dígitos os respectivos símbolos, e a **Evocação Livre**, registo numa folha em branco de todos os símbolos que o examinado consegue recuperar da memória. O outro procedimento, o **CÓDIGO: CÓPIA**, também opcional, é aplicado no final de toda a prova e não apela já à memória, antes requer coordenação visual-motora, pois consiste em copiar directamente os símbolos de uma linha superior para as respectivas células na linha inferior. A aplicação destes procedimentos opcionais procura averiguar se um eventual desempenho deficitário no subteste de Código: Dígito-Símbolo poderá dever-se a dificuldades ao nível da memória ou da motricidade. Este subteste apresenta o interesse de ser a única tarefa de aprendizagem de toda a bateria, a velocidade de desempenho dependendo da facilidade e da qualidade desta aprendizagem, a qual é influenciada pela eficácia da memória visual a curto prazo, pela atenção e pela concentração. Mas há que não esquecer que a velocidade do desempenho grafo-motor também é determinante para a rapidez e execução deste subteste, razão por que este aspecto, associado à velocidade de processamento, pode explicar o rápido e precoce declínio das médias com a idade. De facto, este subteste da *WAIS-III* saturou no factor de Velocidade de Processamento juntamente com o subteste de Pesquisa de Símbolos, quer na amostra total (saturação de .68), quer em todos os níveis etários, com saturações entre .52 e .74, e, como antes assinalado, no nível etário mais avançado, ainda em conjunto com os testes da parte de realização que envolvem limites de tempo (Completamento de Gravuras, Cubos e Disposição de Gravuras). O seu valor como medida de *g* é modesto, saturação de .59; o índice de precisão, neste caso (um teste de velocidade) não um índice de consistência interna mas um índice médio de estabilidade temporal, é de .81.

SUBTESTE DE PESQUISA DE SÍMBOLOS. Introduzido primeiro na *WISC-III* (em 1991), vem a ser acrescentado à *WAIS* na sua terceira edição com o intuito de reforçar a medida da velocidade de processamento. Não é necessário aplicá-lo para o apuramento de QIs mas contribui para o ìndice Factorial de Velocidade de Processamento. É realizado num formulário à parte da Folha de Registo, o Caderno de Respostas da Pesquisa de Símbolos, e foi adaptado da tarefa concebida por Samuel Sternberg em 1966 (citado em Grégoire, 2004) para o estudo do procedimento de verificação da

pertença de um símbolo estímulo a uma série de símbolos, tarefa mais tarde aplicada também ao estudo do controlo da atenção. É composto por 60 séries de símbolos, cada uma precedida de dois símbolos alvo, e a tarefa, que tem um tempo limite de 120 segundos, consiste em decidir se um dos símbolos alvo faz parte da série dada, assinalando a resposta "sim" ou a resposta "não". Apesar de constituir uma medida medianamente robusta da inteligência geral, posto que a saturação em *g* é de apenas .70, terá cumprido o objectivo da sua inclusão na bateria já que satura no factor de Velocidade de Processamento em conjunto com o Código: Dígito-Símbolo, com saturações de .63, na amostra total, e entre .59 e .72, nas amostras etárias, sendo que no grupo de idade mais avançada, como já assinalado, se agrupa ainda com os subtestes de realização em que há limites de tempo de execução (Completamento de Gravuras, Cubos e Disposição de Gravuras). A média dos coeficientes de estabilidade temporal (procedimento aplicado por ser um teste de velocidade) foi, na *WAIS-III,* de .77.

SUBTESTE DE COMPOSIÇÃO DE OBJECTOS. Anteriormente integrado na parte de Realização da bateria, este subteste passou a opcional na última edição, não sendo necessário aplicá-lo para a obtenção de nenhum dos resultados compósitos (não contribui para o apuramento nem dos QIs, nem dos Índices Factoriais). A decisão de passagem a opcional baseou-se nos resultados metrológicos que continuadamente produziu, ao longo das anteriores edições, uma vez que o pequeno número de itens (cinco) não permitia a obtenção de bons índices de consistência interna (na *WAIS-III,* a média dos índices de consistência situou-se em .70), além de que não proporcionava uma medida suficientemente robusta da inteligência geral (saturação em *g* de .62). Assim, a ampliação da extensão do teste tornava-se supérflua, do ponto de vista da medição da inteligência, e inviável, do ponto de vista do tempo de aplicação, uma vez que este teste consumia já parte substancial do tempo total despendido na administração da prova. Ainda assim, o teste foi não só mantido, na qualidade de opcional, como aperfeiçoado nos seus conteúdos, pela substituição de itens menos bons pelas qualidades técnicas ou pela desactualização do conteúdo. O subteste comporta, então, 5 itens de construção de *puzzles* em que as peças são apresentadas separadas, de acordo com disposições padrão, e a tarefa consiste em reuni-las para formar um objecto; note-se que neste subteste, ao contrário do de Cubos, não existe modelo do objecto a construir, e cabe ao próprio examinado identificar qual o objecto em cada item. Decorre daqui o enorme valor clínico deste subteste, já sublinhado por Wechsler, dado que proporciona oportunidade óptima de observação do examinado em situação de resolução de problemas, revelando aspectos conativos e afectivos da actividade cognitiva – estratégias de resolução (planeamento, ensaio e erro, etc.), reacção à frustração, persistência, impulsividade, percepção de auto-eficácia, estilo cognitivo, etc. Por não entrar no apuramento dos resultados compósitos, este subteste foi omitido no estudo factorial da *WAIS-III*; contudo, os resultados obtidos em estudos factoriais anteriores em geral

demonstraram que tende a agrupar-se com Cubos, Completamento de Gravuras e Disposição de Gravuras no factor de Organização Perceptiva (Grégoire, 2004; Kaufman & Lichtemberger, 1999).

A caracterização dos catorze subtestes incluídos na *WAIS-III* desde logo esclarece sobre a elevada qualidade metrológica das medidas que proporcionam: como se viu, os dados provenientes do estudo da precisão, obtidos através do estudo da consistência interna, na maioria dos subtestes (à excepção dos testes de velocidade), situaram-se entre .70 e .93 (mediana .84) e a estrutura factorial identificada corresponde à concepção de inteligência subjacente à prova: todos os subtestes constituem medidas de *g*, ainda que em grau variável (saturações em g entre .57 e .83) e saturam secundariamente em um de quatro factores de grupo, os dois mais importantes, que explicam a maioria da variância dos resultados, com alguma correspondência com a subdivisão da escala em partes Verbal e de Realização (os factores de Compreensão Verbal e de Organização Perceptiva, respectivamente). Importa ainda referir que o manual da terceira edição acrescenta muitos outros resultados empíricos importantes para o esclarecimento das potencialidades metrológicas das medidas, relativos aos subtestes bem como aos resultados compósitos:

- Relativamente ao estudo da **PRECISÃO**, outras técnicas de estimação foram utilizadas, designadamente para estudo da estabilidade temporal (teste-reteste) e do acordo inter-avaliadores, este relevante sobretudo nos subtestes de Vocabulário, Semelhanças e Compreensão, cujas respostas exigem julgamento por parte do avaliador ao serem classificadas com recurso a critérios previamente estabelecidos. Os índices de estabilidade temporal, para o conjunto de todas as idades[13], situaram-se entre .69 (Disposição de Gravuras) e .94 (Informação) e no estudo do acordo inter-avaliadores foram obtidas as seguintes correlações: .95 para o Vocabulário, .93 para as Semelhanças e .91 para a Compreensão. Ao nível dos resultados compósitos, os índices de precisão são, como seria de esperar, ainda mais elevados (na lista que se segue, o primeiro índice diz respeito à consistência interna[14], o segundo à estabilidade temporal[15], e os valores apresentados representam as médias de todos os grupos etários):
  - QI Escala Completa: .98 / .96
  - QI Verbal: .97 / .96
  - QI de Realização: .94 / .91
  - Índice de Compreensão Verbal: .96 / .95

[13] Wechsler, 1997, p.61, Tabela 3.9.
[14] Wechsler, 1997, p.50, Tabela 3.1.
[15] Wechsler, 1997, p.61, Tabela 3.9.

- Índice de Organização Perceptiva: .94 / .88
- Índice de Memória de Trabalho:.93 / .89
- Índice de Velocidade de Processamento: .87 / .89

- Os estudos de **VALIDAÇÃO** cobriram um notável conjunto de técnicas, designadamente, estudos de validação intra-conceito – análises factoriais exploratória e confirmatória – e estudos de validação inter-conceito – estudos de validação convergente e discriminante – correlações com outros testes de inteligência e com outras medidas da cognição, atenção e concentração, memória, linguagem, velocidade e destreza motora, processamento espacial, funcionamento executivo – e estudos com grupos especiais – com perturbações neurológicas (Alzheimer e Huntington), perturbações de dependência do álcool (Korsakoff), perturbações neuropsiquiátricas (esquizofrenia), perturbações do desenvolvimento psicoeducacional (perturbações da aprendizagem e deficiência mental) e deficiências visual e auditiva (Wechsler, 1997b, pp.75-180, Capítulo 4). De todo este conjunto de evidências de validade, retêm-se aqui as que se mostram mais pertinentes do ponto de vista dos propósitos da presente investigação, a análise factorial e a correlação com outras medidas da inteligência.
  - Validação intra-conceito:
    - Análise Factorial Exploratória. Foi aplicado o método dos eixos principais e rotação oblíqua com retenção de quatro factores (tomando como critério para a selecção de factores a literatura anterior e a estrutura factorial da *WISC-III*) (N=1250): os quatro factores são, por ordem da proporção da variância que explicam, Compreensão Verbal, Organização Perceptiva, Memória de Trabalho e Velocidade de Processamento e têm correlações entre si que oscilam entre .60 e .77 (percentagem de variância comum entre os factores igual ou inferior a 60%). A adição dos subtestes de Ordenação de Letras e Números e de Pesquisa de Símbolos tornou mais robustos os factores respectivos, Memória de Trabalho e Velocidade de Processamento. De uma maneira geral, o estudo factorial exploratório nas amostras etárias confirmou o mesmo padrão de resultados, excepto para o grupo etário mais idoso em que as saturações nos factores de Organização Perceptiva e de Velocidade de Processamento são menos claras (subtestes com tempo limite, que em idades inferiores

saturam no factor de Organização Perceptiva, nesta faixa de idade saturam no factor Velocidade de Processamento).

- Análise Factorial Confirmatória. Testou o ajustamento dos dados a cinco modelos, definidos por um a cinco factores, respectivamente, tendo confirmado que o modelo que melhor se ajusta aos dados é o de quatro factores (equivalentes aos da análise factorial exploratória), quer na amostra total (N=1250) quer nas subamostras etárias (n entre 200 e 300) (embora o ajustamento a cinco factores – Aritmética isolada num factor de Aptidão Numérica ou Quantitativa – apresente índices de ajustamento próximos, e até ligeiramente melhores no grupo etário mais avançado). Para a amostra total, os índices de ajustamento do modelo de quatro factores foram:
  - χ^2/gl=4.1, AGFI=.954, RMSR=.221, TLI=.82 (comparação com modelo de factor geral)

- Validação inter-conceito:
  - Correlações com outras medidas da inteligência (validação convergente) (Wechsler, 1997b, pp.78-91):
    - *WAIS-R* (N=192): para os subtestes, correlações entre .50 (Completamento de Gravuras) e .90 (Vocabulário); para os QIV, QIR e QIEC, respectivamente, .94, .86 e .93.
    - *WISC-III* (N=184): para os subtestes, correlações entre .31 (Disposição de Gravuras) e .83 (Vocabulário); para os QIV, QIR e QIEC, respectivamente, .88, .78 e .88; e para os ICV, IOP, IMT e IVP, respectivamente, .87, .74, .80 e .79.
    - *Stanford-Binet Intelligence Scale – Fourth Edition* (N=26): correlações entre o resultado total da *S-B IV* e os QIV, QIR e QIEC, respectivamente, .78, .89, .88; e correlações com os ICV, IOP e IVP, respectivamente, .85, .86 e .07.
    - *Standard Progressive Matrices (SPM)* (N=26): correlações do resultado total do teste *SPM* (1976) com os QIV, QIR e QIEC, respectivamente, .49, .79 e .64; com ICV, IOP e IVP, respectivamente, .55, .65 e .25; subteste *WAIS-III* mais correlacionado com total *SPM*: Matrizes, .81.

- Estas e outras evidências de validade, cujo tratamento seria demasiado extenso para inclusão exaustiva no presente ponto do trabalho, evidenciam em geral o elevado valor metrológico desta prova, fundamentam a interpretação dos seus resultados como medidas da inteligência, tal como Wechsler a concebe, e legitimam a sua utilização em contextos em que a avaliação da inteligência se mostra pertinente, em particular nos contextos educacional e clínico.[16]

A robustez técnica desta prova, aliada ao carácter compósito e estrutural (hierárquico), com correspondência nos modelos diferenciais mais consensuais, fundamentam a sua escolha como representante do paradigma diferencial de avaliação da inteligência, no quadro do presente estudo.

2) **TESTE TRIÁRQUICO DE APTIDÕES DE STERNBERG - REVISTO: NÍVEL H *[STAT-R (H)]***

O *STAT-R (H)* é uma técnica de avaliação da inteligência construída por Sternberg em 1993, como operacionalização da sua Teoria Triárquica da Inteligência Humana, e permanece até hoje em fase de desenvolvimento experimental[17]. A edição revista do nível destinado ao ensino superior e aos adultos, o Nível H, data de 2002 (Sternberg, 2002d,e), e foi traduzida e adaptada para Portugal no âmbito da presente investigação.

Na comparação com a *WAIS-III*, desde logo há que assinalar tratar-se de uma prova muito mais incipiente e tentativa, com um corpo de investigação empírica muito menos extenso e sólido e enquadrada conceptualmente por um modelo, ele próprio ainda sob forte questionamento e em permanente aperfeiçoamento[18]. Como acontece com os testes em fase experimental, ainda não dispõe de manual, onde se encontrem reunidas as informações pertinentes sobre a prova, a sua aplicação, cotação e interpretação de resultados, sendo o material composto por apenas três elementos: o caderno de teste, a folha de respostas e as instruções de aplicação. Entre a versão original (1993) e a versão revista (2002) do *STAT(H)*, não houve alterações na estrutura da prova, apenas no seu conteúdo: cada parte do teste passou de quatro para cinco itens, aumentando a extensão total de 36 para 45 itens e houve reformulações e substituições, principalmente ao nível dos itens mas também das instruções.

A prova é composta por duas secções, que podem ser aplicadas separadamente, as quais se distinguem essencialmente em relação ao formato de resposta – itens de resposta por escolha múltipla,

[16] Pela natureza da presente investigação, os resultados da análise factorial, exploratória e confirmatória, serão considerados em algum detalhe no CAPÍTULO 8, na discussão dos resultados do Estudo Principal.
[17] Embora tenham sido iniciadas negociações entre o autor e a *Psychological Corporation*, hoje pertencente à Harcourt Assessnment (a editora das escalas de Wechsler), o teste não chegou a alcançar a fase de publicação e continua a ser utilizado exclusivamente em contexto de investigação (ver por exemplo Sternberg et al., 2004, 2006).
[18] Como se verá, este desnível entre os instrumentos, quanto à solidez dos seus fundamentos e ao nível de operacionalização, constituiu uma das principais dificuldades metodológicas desta investigação.

na primeira, e itens ensaio ou de resposta aberta, na segunda. Os itens do primeiro tipo propõem quatro alternativas de resposta e a cotação é objectiva, efectuada com recurso a uma chave de cotação, fornecida pelo autor juntamente com o material; os itens do segundo tipo requerem julgamento e ponderação por parte do examinador, recorrendo a um conjunto de critérios baseados na análise da forma e do conteúdo das produções escritas dos examinados. Apesar da menor objectividade de classificação, estes itens dão lugar a uma mais livre e espontânea expressão, o que é sem dúvida importante sobretudo na avaliação de áreas de funcionamento que envolvem o pensamento divergente (inteligência criativa). O teste foi concebido para aplicação colectiva com tempo limite – cinco minutos para cada parte da primeira secção, quinze minutos para cada ensaio da segunda secção – embora esteja prevista a possibilidade de prolongamento da aplicação, na primeira secção, até ao máximo de oito minutos, quando os examinados não consigam terminar todos os itens dentro do limite inicialmente estabelecido.

A TABELA 4.3 representa a estrutura original do *STAT-R (H)*. A primeira secção comporta nove partes que resultam do cruzamento das três formas de inteligência postuladas pela teoria, ou "domínios de processamento mental" – Inteligência Analítica, Inteligência Prática e Inteligência Criativa –, com três tipos de conteúdo, ou "áreas de representação mental da informação" – Verbal, Quantitativa e Figurativa. A segunda Secção do teste corresponde à Parte 10 e contém três Itens Ensaio, um relativo a cada forma de inteligência – Ensaio Analítico, Ensaio Prático e Ensaio Criativo.

Nos primeiros esforços de operacionalização da Teoria Triárquica, Sternberg concebeu o teste com uma estrutura aproximada da que ele veio a adquirir, mas com algumas diferenças a assinalar (Sternberg, 1991): primeiro, as três partes hoje designadas de Analíticas eram equacionadas como operacionalização da Subteoria Componencial, pelo que supostamente avaliavam "o aspecto componencial da inteligência"; em conformidade, os testes designavam-se de Componencial:Verbal, Componencial:Quantitativo e Componencial:Figurativo, sendo o seu conteúdo semelhante ao das actuais Partes 1, 2 e 3; segundo, a operacionalização da Subteoria Experiencial desdobrava-se em duas categorias diferenciadas de testes, três para avaliar a capacidade para lidar eficientemente com situações novas – Lidar com a Novidade:Verbal, Lidar com a Novidade:Quantitativo e Lidar com a Novidade:Figurativo – e três testes de velocidade, para testar a capacidade de automatização – Automatização:Verbal (letras), Automatização:Quantitativo (números) e Automatização:Figurativo (figuras geométricas) – estes três últimos testes vieram a ser abandonados na construção do *STAT*; por fim, não toda a Subteoria Contextual, mas apenas uma das funções que nela são identificadas – a Adaptação – era operacionalizada em três testes criados para avaliar o Aspecto Prático da Inteligência:Verbal, o Aspecto Prático da Inteligência:Quantitativo e o Aspecto Prático da Inteligência:Figurativo. Esta presença no *STAT* apenas da função adaptativa da inteligência, com

TABELA 4.3
Estrutura do STAT-R (H)

FORMAS DE INTELIGÊNCIA **Domínios de Processamento Mental**	**CONTEÚDOS** **Áreas de Representação Mental da Informação**			***PARTE 10***
	VERBAL (V)	**QUANTITATIVA** (Q)	**FIGURATIVA** (F)	
INTELIGÊNCIA ANALÍTICA (IA)	***PARTE 1*** Analítica Verbal (AV)	***PARTE 2*** Analítica Quantitativa (AQ)	***PARTE 3*** Analítica Figurativa (AF)	*Ensaio Analítico*
INTELIGÊNCIA PRÁTICA (IP)	***PARTE 4*** Prática Verbal (PV)	***PARTE 5*** Prática Quantitativa (PQ)	***PARTE 6*** Prática Figurativa (PF)	*Ensaio Prático*
INTELIGÊNCIA CRIATIVA (IC)	***PARTE 7*** Criativa Verbal (CV)	***PARTE 8*** Criativa Quantitativa (CQ)	***PARTE 9*** Criativa Figurativa (CF)	*Ensaio Criativo*

omissão da Modelação e da Selecção, caracteriza ainda hoje a última edição do teste. Deste ponto de vista, a estrutura actual do *STAT* contém três testes que representam a Subteoria Componencial – Partes 1, 2 e 3 (Analíticas) – três testes que representam uma das funções da Subteoria Contextual, a Adaptação – Partes 4, 5 e 6 (Práticas) – e três testes que representam apenas uma parte da Subteoria Experiencial, as capacidades próximas do pólo "novidade", no contínuo novidade-automatização - Partes 7, 8 e 9 (Criativas).

Apesar da redução da extensão do teste desde a primeira tentativa de operacionalização da teoria, de 12 para 9 subtestes (pela eliminação dos testes de Automatização), a duração total da aplicação do *STAT-R (H)* (2002) excede as duas horas, demorando cerca de 90 minutos a aplicação apenas da primeira secção – cerca de 45 minutos de execução e 45 minutos gastos com as instruções, que incluem sempre dois itens exemplo e são administradas separadamente para cada parte. Assim,

apesar de se reconhecer as potencialidades da Parte 10, no âmbito da presente investigação, por razões de ordem prática, optou-se por utilizar apenas a primeira secção do teste: primeiro, porque no Estudo Principal, em que se previa aplicar as duas técnicas, *WAIS-III* e *STAT-R (H)*, na mesma amostra, o tempo de aplicação excederia três horas e meia, o que dificultaria o recrutamento de participantes adultos; segundo, porque a classificação das respostas da Parte 10 deve ser preferencialmente efectuada por pelo menos dois juízes independentes, o que representaria uma sobrecarga na organização do trabalho experimental e no investimento em recursos humanos. Por fim, porque os estudos factoriais do *STAT-R (H)* omitem em geral os Itens Ensaio, talvez por este tipo de estudos exigir um número elevado de participantes, dificultando a aplicação de técnicas de classificação e cotação morosas, como as que envolvem a análise de conteúdo de textos.

As variáveis do *STAT-R (H)* retidas para análise são, assim, os resultados das nove partes, bem como os seis resultados compósitos representando ou os domínios de processamento – Inteligência Analítica (IA), Inteligência Prática (IP), Inteligência Criativa (IC) – ou as áreas de conteúdo – Verbal (V), Quantitativa (Q) e Figurativa (F). Cada um destes resultados é obtido a partir do somatório dos resultados de três partes, 15 itens, ou com conteúdos diversos (V + Q + F), ou envolvendo áreas de processamento diversas (IA + IP + IC); os resultados Analítico, Prático e Criativo fundamentam-se na Teoria Triárquica; os resultados Verbal, Quantitativo e Figurativo enquadram-se na perspectiva multifactorial e pode admitir-se remeterem para três aptidões clássicas: aptidão verbal, aptidão numérica e aptidão espacial (ver CAPÍTULO 1, p.46). O seu interesse na presente investigação liga-se à possibilidade de teste de estruturas alternativas para a organização das mesmas variáveis cognitivas.

Embora por vezes nos trabalhos com o *STAT-R (H)* seja feita referência ao apuramento de um resultado total (Carrasco, 2000; Sternberg, Castejón, Prieto, Hautamäki & Grigorenko, 2001; Sternberg, Prieto & Castejón, 2000; Sternberg & The Rainbow Project Collaborators, 2006), quer a teoria subjacente ao teste, quer a sua própria estrutura, conferem a um tal resultado um estatuto pouco fundamentado. Com efeito, Sternberg, à semelhança dos outros autores sistémicos, recusa a inteligência como conceito holístico e sublinha que o que define a inteligência funcional não é a "quantidade" de inteligência (que poderia ser estimada a partir do somatório de todos os itens ou resultado total da bateria) mas o "equilíbrio" na gestão das diferentes formas de inteligência (mais a integração do que a adição entre as várias formas de inteligência); por outro lado, a heterogeneidade de áreas de processamento e de conteúdo contempladas pelo *STAT-R (H)* tornaria um tal resultado internamente pouco consistente, tanto mais que, ao contrário de Wechsler, Sternberg não se propõe medir a inteligência de várias maneiras, mas sim medir "diferentes formas" de inteligência. Ao contrário dos seis resultados compósitos parciais que, apesar de alguma heterogeneidade decorrente da reunião de três partes do teste, representam sempre um domínio ou uma área homogénea, ou de

processamento, ou de conteúdo, o eventual resultado total obtido a partir do somatório de todos os itens não encontra qualquer tipo de fundamento, pelo menos na fase actual de desenvolvimento da investigação com o teste, razão por que não será objecto de tratamento na presente investigação.

Caracterizado o *STAT-R (H)* na generalidade, importa passar à apresentação das nove partes do ponto de vista do conteúdo (Sternberg, 2002d,e), racional subjacente à construção (Sternberg, 1985a, 1991, 1992) e dados metrológicos disponíveis – coeficiente de Kuder-Richardson-20, correlação item-parte (com o item omitido do total da respectiva parte) e análise factorial confirmatória – obtidos com a primeira versão do *STAT (H)* (1993) (Sternberg, 1997d; Sternberg, Castejón, Prieto, Hautamäki & Grigorenko, 2001; Sternberg & Clinkenbeard, 1995; Sternberg, Ferrari, Clinkenbeard & Grigorenko, 1996; Sternberg, Grigorenko, Ferrari & Clinkenbeard, 1999; Sternberg, Prieto & Castejón, 2000) e com a segunda versão, o *STAT-R (H)* (2002) (Sternberg & The Rainbow Project Collaborators, 2006).

PARTE 1: ANALÍTICA VERBAL: Nesta prova a tarefa do examinado consiste em extrair vocabulário do contexto Cada item consta de um pequeno texto que contém uma palavra sem significado ou neologismo; as alternativas de resposta identificam quatro possíveis significados dessa palavra, dos quais apenas um se aplica ao contexto. Embora esta parte procure avaliar um domínio de processamento classicamente presente nos testes de inteligência – a compreensão verbal ou inteligência analítica aplicada a conteúdos verbais – adopta um formato inovador que se centra no processo de aquisição de vocabulário (processo de aprendizagem), não no produto ou conhecimento do vocabulário adquirido no passado (produto da aprendizagem), como acontece nos testes tradicionais de vocabulário (Vocabulário da *WAIS-III,* testes de sinónimos e de antónimos, etc.). A justificação lógica para uma tal opção reside na verificação empírica (Sternberg, 1987) de que a maioria do vocabulário é aprendido informalmente, por inferência, a partir de pistas contextuais, a sua avaliação permitindo a aproximação a uma competência mais fundamental, a capacidade de recolher informação a partir de contextos relevantes. Os índices de consistência interna disponíveis na literatura, obtidos em três amostras com a primeira versão do *STAT (H)* (1993)[19] que, recorde-se, continha apenas quatro itens em cada parte, foram de .32 (para a versão original do teste), .01 (para a versão finlandesa) e .25 (para a versão espanhola). As correlações entre cada item e a parte Analítica-Verbal (soma de 3 itens, omisso cada item em estudo) variaram entre .15 e .20 (versão original), -.03 e .04 (amostra finlandesa) e .05 e .19 (amostra espanhola); as correlações entre os itens e toda a Parte Analítica (11 itens, omisso o respectivo item) variaram entre .13 e .24 (versão original), -.06 e .11 (amostra finlandesa) e -.03 e .20

[19] Os índices relativos às nove partes do *STAT (H)* (1993) são apresentados num único artigo (Sternberg, Castejón, Prieto, Hautamäki & Grigorenko, 2001, p.6, Tabela 3), o mesmo onde se recolheram as correlações item-parte (p.7, Tabela 4) e os dados relativos à análise factorial confirmatória (pp.11-13 e Tabela 6). As características das três amostras em que foram obtidos estes resultados são as seguintes: amostra americana: 326 estudantes, 14 a 18 anos; amostra finlandesa: 2712 estudantes do 6º ano de escolaridade, com 12 a 13 anos de idade; amostra espanhola, 240 estudantes do 10º ano de escolaridade com média etária de 16.4 anos.

(amostra espanhola). Tomando o modelo que na análise factorial confirmatória obteve melhores índices de ajustamento[20], as saturações (parâmetros estimados) da Parte Analítica-Verbal no factor de segunda ordem Analítico foram, para a versão original do teste .42, para a versão finlandesa -.23 e para a versão espanhola .52.

PARTE 2: ANALÍTICA QUANTITATIVA: Trata-se de um teste comum de séries numéricas: em cada item a tarefa consiste em indicar, de entre as quatro alternativas de resposta, qual a que completa a sequência numérica apresentada. As sequências variam quanto à extensão e quanto ao número e complexidade das regras de ordenação subjacentes. Este tipo de itens mede o raciocínio indutivo no domínio numérico pelo que a fundamentação lógica subjacente à inclusão deste tipo de tarefa no *STAT* provém da investigação previamente conduzida pelo autor e colaboradores para estudo do raciocínio indutivo (Sternberg, 1977, 1985a; Sternberg & Gardner, 1983). A evidente semelhança entre o conteúdo e a tarefa desta parte do teste e alguns testes comuns de raciocínio numérico deve-se a que a inteligência analítica é a forma de inteligência da Teoria Triárquica que, de acordo com Sternberg, se encontra representada nas concepções e nos métodos de medida da inteligência tradicionais. Os índices de KR-20 obtidos no estudo do *STAT (H)* (1993) (4 itens) foram de .64 (USA), .17 (Finlândia) e .53 (Espanha), as correlações dos itens com a parte Analítica-Quantitativa situaram-se entre .33 e .45 (USA), -.01 e .17 (Finlândia) e .24 e .37 (Espanha) e as correlações dos itens desta parte com toda a Parte Analítica situaram-se entre .38 e .43 (USA), -.01 e .13 (Finlândia) e .26 e .29 (Espanha). Na análise factorial confirmatória, os parâmetros estimados no modelo mais ajustado, para a parte Analítica-Quantitativa - saturações no factor de segunda ordem Analítico – foram .59 (USA), .26 (Finlândia) e .37 (Espanha).

PARTE 3: ANALÍTICA FIGURATIVA: É uma prova clássica de matrizes figurativas, neste caso composta por figuras geométricas, e a tarefa consiste em preencher, com uma das alternativas de resposta, a célula de cada matriz que se encontra em branco. Trata-se de um teste clássico de analogias figurativas que envolve, à semelhança do anterior, o raciocínio indutivo, pelo que se fundamenta na mesma literatura proveniente da investigação do autor e colaboradores nas décadas de 70 e 80 (Sternberg, 1977, 1985a; Sternberg & Gardner, 1983). E, tal como na parte anterior, também nesta a razão para a semelhança do seu conteúdo com testes clássicos, neste caso de matrizes, deve-

[20] De acordo com Sternberg, Castejón, Prieto, Hautamäki & Grigorenko (2001) o modelo derivado da Teoria Triárquica, – partindo da análise ao nível dos itens, e pela aplicação do programa *LISREL 8.12*, nove factores de primeira ordem, correspondentes às nove partes do teste, e três factores oblíquos de segunda ordem, Analítico, Prático e Criativo – é o que melhor se ajusta à estrutura dos dados, em especial se forem admitidas correlações entre os erros ("Modelo 6-ce, *correlated errors*") [alguns índices de ajustamento: χ^2/gl=2.75, GFI=.91, RMSEA=.033, CFI=.83 e RFI=.71 (amostra Finlândia+USA); χ^2/gl=2.21, GFI=.99, RMSEA=.027, CFI=.86 e RFI=.74 (amostra Finlândia+Espanha) e χ^2/gl=0.52, GFI=1.00, RMSEA=.000, CFI=1.0 e RFI=.57 (amostra USA+Espanha)]. As correlações entre os factores de segunda ordem são: Analítico-Prático .93 (a 95% de probabilidade, intervalo de confiança .89-.97); Analítico-Criativo .85 (intervalo de confiança .81-.89) e Prático-Criativo .72 (intervalo de confiança .68-.76).

se a que estes, na óptica de Sternberg, medem essencialmente a inteligência analítica tal como a Teoria Triárquica a concebe. Os índices de consistência interna foram neste caso de .39, .40 e .36, respectivamente para as versões original, finlandesa e espanhola, e as correlações dos itens com a Parte Analítica-Figurativa situaram-se, pela mesma ordem das amostras, entre .15 e .25, entre .15 e .25 e entre .15 e .23; as correlações dos itens com toda a Parte Analítica localizaram-se, ainda pela mesma ordem das amostras, entre .25 e .28, entre .15 e .21 e entre .12 e .24. Tomando o modelo mais ajustado na análise factorial confirmatória, as saturações da parte Analítica-Figurativa com o factor Analítico foram de .43 (USA), .44 (Finlândia) e .07 (Espanha).

PARTE 4: PRÁTICA VERBAL: Em cada item é apresentada uma breve descrição de uma situação prática da vida quotidiana de um/uma estudante, na qual há que tomar uma decisão. A tarefa consiste em optar por uma entre quatro alternativas de cursos de acção possíveis, tendo em vista a melhor resolução da situação dada. Talvez no intuito de sublinhar o carácter verbal e prático da tarefa, o examinado é convidado a imaginar que escreve na coluna de "consultório" de um jornal para estudantes, e que a alternativa escolhida seria a sugestão ou conselho que daria nessa coluna a um estudante que lhe tivesse colocado o problema descrito. Consiste num teste de raciocínio inferencial aplicado a situações de ordem prática e envolve o raciocínio informal, ou seja, o tipo de raciocínio que utilizaria caso enfrentasse a situação na sua vida quotidiana e tivesse de lhe responder. O fundamento lógico da inclusão deste tipo de teste no *STAT* remete para a investigação de Sternberg e colaboradores no domínio da inteligência prática, e, em particular, sobre o "conhecimento tácito" (por exemplo, Sternberg & Wagner, 1993; Wagner & Sternberg, 1985, 1986): está em causa a aplicação do conhecimento tácito à resolução de problemas que se colocam na relação do indivíduo com o contexto (Subteoria Contextual) quando tem em vista a adaptação a esse contexto (não a sua modelação ou a selecção de um novo meio). Os índices de consistência interna KR-20 obtidos nesta parte do teste foram .04 para a versão original, .16 para a amostra finlandesa e .12 para a amostra espanhola. As correlações entre os itens e o total da Parte Prática-Verbal situaram-se entre .01 e .12 (EUA), entre .04 e .11 (Finlândia) e entre -.01 e .12 (Espanha); e as correlações entre os itens e o total da parte Analítica localizaram-se entre .07 e .14 (EUA), entre .03 e .12 (Finlândia) e entre .07 e .14 (Espanha). Quanto à estrutura factorial, as saturações estimadas entre esta parte e o factor de Inteligência Prática do modelo mais ajustado aos dados foram de .24, .13 e .05, respectivamente nos EUA, na Finlândia e em Espanha.

PARTE 5: PRÁTICA QUANTITATIVA: Compõe-se de problemas quantitativos envolvidos na resolução de tarefas da vida quotidiana, como seguir as indicações de receitas culinárias, calcular médias de classificações escolares ou raciocinar sobre fusos horários. À semelhança do teste anterior, envolve o raciocínio tal como é informalmente aplicado no dia a dia quando se enfrentam situações e

resolvem problemas com dados quantitativos. Como tal, o fundamento lógico para a inclusão desta parte da bateria encontra-se nos mesmos conceitos de inteligência prática e de conhecimento tácito, aplicados agora a conteúdos quantitativos (Sternberg & Wagner, 1993; Wagner & Sternberg, 1985, 1986). Os coeficientes KR-20 desta parte do teste foram de .55 na amostra americana, .20 na finlandesa e .30 na espanhola. As correlações entre os itens e o total da Parte Prática-Quantitativa situaram-se nas mesmas amostras, e pela ordem respectiva, entre .32 e .36, entre .07 e .13 e entre .11 e .21; e as correlações com a Inteligência Prática, respectivamente, entre .27 e .35, entre .01 e .14 e entre .12 e .27. Ao retomar a estrutura factorial que na análise confirmatória mais se aproximou da estrutura dos dados, verificou-se que as saturações estimadas desta parte no factor de Inteligência Prática foram de .47, na versão original do teste, .03 na versão finlandesa e .54 na versão espanhola.

PARTE 6: PRÁTICA FIGURATIVA: Trata-se de uma prova de orientação em mapas: em cada item há que encontrar um caminho entre dois pontos através de um mapa, respeitando determinados condicionalismos que são parte do problema. No último item a tarefa é um pouco diferente, consiste em destinar o lugar que deve ocupar determinado convidado no diagrama da mesa de um jantar, em função de um conjunto de condições que o problema estabelece. Embora as situações de ordem prática que há que resolver envolvam nesta parte do teste conteúdos figurativos (mapas ou diagramas), a fundamentação lógica do teste reside ainda nos trabalhos sobre a inteligência prática e o conhecimento tácito desenvolvidos pelo autor e seus colaboradores, aqui aplicados a conteúdos figurativos (Sternberg & Wagner, 1993; Wagner & Sternberg, 1985, 1986). A inspecção dos índices de consistência interna das versões americana, finlandesa e espanhola do teste revela índices, respectivamente, de .14, .15 e .09. As correlações dos itens com a Parte Prática-Figurativa situaram-se entre .00 e .11 , entre .03 e .10 e -.02 e .09, respectivamente nos EUA. na Finlândia e em Espanha, e as correlações com o resultado Prático total entre .11 e .21, entre .03 e .16 e entre .12 e .18, nas mesmas amostras. Na estrutura factorial mais ajustada aos dados, na análise confirmatória, as saturações estimadas para esta parte do teste no factor de Inteligência Prática foram de .16 (EUA), .11 (Finlândia) e .40 (Espanha).

PARTE 7: CRIATIVA VERBAL: Envolve a resolução de analogias verbais aparentemente vulgares mas que devem ser respondidas assumindo uma premissa contrafactual (como, "o dinheiro cai das árvores"). Por vezes essa premissa é relevante para a resolução da analogia, outras vezes não é, cabendo ao examinado começar por julgar a pertinência da premissa para a resolução da analogia. Este tipo de teste foi integrado no *STAT-R (H)* para avaliar a facilidade de resolução de problemas relativamente novos, o que remete para a noção, central na Subteoria Experiencial, de que a inteligência é melhor avaliada em determinados pontos do contínuo que se estabelece entre a total novidade e a automatização, aqui num ponto que se situa próximo mas não no extremo da absoluta

novidade. Nas três partes criativas do teste (Partes 7 a 9), não está em causa a produção divergente, isto é, a criação de novos produtos, mas a capacidade de rapidamente responder com eficácia perante uma situação relativamente nova. Nesta parte, a situação nova é introduzida, precisamente, pelas premissas contrafactuais. A fundamentação para a construção desta parte do *STAT* remete para as investigações realizadas por Sternberg e Gastel (1989a,b) que mostraram correlações entre os tempos de resposta a premissas factuais e contrafactuais e os resultados de testes de inteligência fluida, sugerindo que a avaliação da capacidade para lidar com a novidade constitui uma possível forma de medir da inteligência. Nesta parte do teste, os coeficientes de consistência interna foram de .37 para a forma original do teste, .01 para a forma finlandesa e -.22 para a forma espanhola. As correlações entre os itens e a Parte Criativa-Verbal a que pertencem situaram-se entre .09 e .25, entre .01 e .05 e entre -.30 e .11, respectivamente nas amostras americana, finlandesa e espanhola, e para as mesmas amostras e pela mesma ordem, as correlações com o resultado total Criativo situaram-se entre .16 e .39, entre .03 e .15 e entre .00 e .26.

PARTE 8: CRIATIVA QUANTITATIVA: Propõe três operações aritméticas novas – *graf, flix, e trup* – e a tarefa consiste em resolver expressões numéricas aplicando essas operações. Por exemplo, *flix* é uma operação em que a manipulação numérica a efectuar varia em função da relação entre os dados do problema (a>b, b>a ou a=b). O formato deste teste modificou-se entre as primeiras tentativas de operacionalização e a versão de 2002: inicialmente, o teste era constituído por matrizes numéricas em que alguns algarismos eram substituídos por símbolos, consistindo a tarefa na identificação dos algarismos que teriam de substituir esses símbolos, tendo em vista a resolução da matriz. O fundamento lógico da inclusão deste tipo de tarefa remete para a concepção da criatividade no quadro da Teoria do Investimento (Sternberg & Lubart, 1991b). O abandono deste formato deveu-se, possivelmente, ao reconhecimento de que a tarefa de substituição não é tão pouco familiar quanto seria necessário para esta parte do teste, mais exigindo o raciocínio analítico do que a capacidade para lidar com a novidade (as equações matemáticas utilizam vulgarmente letras em substituição de dados desconhecidos). A opção pelas novas operações matemáticas pode ser tomada como exigindo em maior grau a capacidade para lidar com a novidade, no entanto a resolução dos itens acima de tudo envolve a aplicação de condições na resolução de expressões numéricas simples, um tipo de exercício ainda relativamente comum na aprendizagem da matemática. Os coeficientes de consistência interna do *STAT- (H)* de 1993 (que continha já itens do tipo dos actuais) foram de .53 (USA), .49 (Finlândia) e .74 (Espanha). As correlações entre os itens e a parte do teste respectiva, Criativa-Quantitativa, localizaram-se entre .23 e .46 na amostra americana, entre .14 e .39, na amostra finlandesa e entre .21 e .71, na amostra espanhola e as correlações com o total Criativo alcançaram valores entre .27 e .33, entre .20 e .34 e entre .29 e .63, respectivamente nas mesmas amostras. As saturações estimadas,

para o modelo mais ajustado aos dados, foram nesta parte do teste de .71, .41 e .75 no factor de Inteligência Criativa, respectivamente nas amostras americana, finlandesa e espanhola.

PARTE 9: CRIATIVA FIGURATIVA: Consiste num teste de séries figurativas, com conteúdo geométrico, em que a regra para resolução em cada item não é extraída na própria série a completar, que só tem um elemento, mas em outra série cujo conteúdo não é idêntico. A capacidade para lidar com a novidade é chamada a intervir porque o completamento das séries é feito num domínio diferente daquele em que foi inferida a regra implicando, consequentemente, um processo de transposição (*mapping*); a fundamentação lógica para a construção desta parte provém de investigação desenvolvida por Sternberg e Gardner (1983). Nesta parte do teste, os coeficientes de KR-20 encontrados nas três amostras, americana, finlandesa e espanhola, foram .27, .26 e .56, respectivamente; as correlações entre os itens e a parte Criativa-Figurativa, situaram-se entre .07 e .19 (USA), entre .07 e .22 (Finlândia) e .29 e .39 (Espanha) e as correlações entre os itens e o total Criativo situaram-se, nas mesmas amostras, entre .11 e .30, entre .12 e .25 e entre .30 e .47. Na análise factorial confirmatória os parâmetros (ou saturações no factor de Inteligência Criativa) estimados para esta parte do teste foram de .28, na amostra americana, .29, na amostra finlandesa e .53, na amostra espanhola.

Algumas observações se impõem desde já quando perante o conjunto de dados que acaba de se apresentar: primeiro, os níveis de consistência interna são regra geral muito baixos, assumindo valores muito inferiores aos valores mínimos aceitáveis para a utilização de um teste em avaliação psicológica ou mesmo para a sua aplicação em investigação. O número de itens reduzido de cada parte do teste é apontado como justificação para o nível excessivamente baixo dos coeficientes de consistência interna (Sternberg, Castejón, Prieto, Hautamäki & Grigorenko, 2001) mas há que assinalar que, apesar de o número de itens ser sempre o mesmo (na versão de 1993, quatro em cada parte) há partes em que o nível de consistência interna atinge valores mais aceitáveis (designadamente, a Parte 8: .53, .49 e .74, nos três países a que se refere o estudo). Em segundo lugar, as correlações dos itens com os totais parciais (cada parte) ou globais (cada forma de inteligência) mostram a maior robustez claramente da versão original do teste (à excepção da Parte Prática-Verbal) do que das suas traduções, especialmente da tradução finlandesa, em que se registou uma quantidade muito significativa de itens com correlações próximas de 0 [21]. A Parte 8, Criativa-Quantitativa, é a que regista, de forma consistente em todas as amostras, correlações item-parte e item-total mais elevadas, ao que, naturalmente, correspondem os referidos coeficientes de consistência interna mais altos. Finalmente, os dados da análise factorial confirmatória estão longe de fornecer uma descrição estrutural sólida das

[21] A estes resultados talvez não tenha sido alheio o facto de a amostra finlandesa ter idade muito inferior àquela para que o teste foi construído (12-13 anos).

medidas proporcionadas pelo teste, tendo mesmo desencadeado acesas criticas dirigidas às metodologias, consideradas pouco consistentes, e às conclusões de Sternberg e colaboradores, consideradas algo forçadas ou, no mínimo, precipitadas (Brody, 2003a; Gottfredson, 2003a) (este assunto será retomado no CAPÍTULO 9, a propósito da discussão de resultados do presente estudo).

Passando aos seis resultados compósitos, os três baseados na Teoria Triárquica (domínios de processamento) – Inteligências Analítica, Prática e Criativa – e os três que remetem para uma concepção multifactorial clássica (áreas de representação mental da informação) – Áreas Verbal, Quantitativa e Figurativa – os dados metrológicos disponíveis são melhores, ainda que situando-se também em geral aquém dos níveis considerados desejáveis, sobretudo para utilização da prova em contexto de avaliação psicológica:

- Quanto à **PRECISÃO**, importa referir os coeficientes de consistência interna para cada uma das seis variáveis globais (compósitas), em diferentes amostras e com diferentes versões dos instrumentos (o que, sublinhe-se, exige prudência na comparação de resultados):
  - nas amostras internacionais do estudo que tem vindo a ser citado (Sternberg, Castejón, Prieto, Hautamäki & Grigorenko, 2001, p.6), os índices de consistência do *STAT (H)* (1993), para os resultados globais (12 itens)[22], foram:
    - Amostra Americana: I.Analítica .54, I. Prática .47 e I. Criativa .57
    - Amostra Finlandesa: I.Analítica .32, I. Prática .28 e I. Criativa .46
    - Amostra Espanhola: I.Analítica .52, I. Prática .42 e I. Criativa .70
    - Amostra Americana: Verbal .51, Quantitativa .73 e Figurativa .57
    - Amostra Finlandesa: Verbal ..20, Quantitativa .43 e Figurativa .46
    - Amostra Espanhola: Verbal .25, Quantitativa .65 e Figurativa .61
  - Num outro artigo, Sternberg, Prieto e Castejón (2000, p.644) apresentam os dados provenientes da mesma amostra espanhola do estudo anterior (coeficientes alfa de Cronbach), mas que são surpreendentemente diferentes dos coeficientes de consistência interna antes citados:
    - I.Analítica .60, I. Prática .60 e I. Criativa .75 [23]
  - Na subamostra americana (N=199, 14-18 anos) retirada da amostra antes referida para participar no programa de Verão de Introdução à Psicologia

[22] O artigo fornece também índices de consistência interna para o conjunto de todo o teste (36 itens) – .82 para a amostra americana, .67 para a amostra finlandesa e .82 para a amostra espanhola – apesar de, por razões conceptuais, este resultado ser pouco fundamentado (ver comentário pp.281-282).

[23] Consistência interna do resultado global de todo o teste (36 itens): .80.

(estudo referido adiante, a propósito da validação externa – Sternberg, Grigorenko, Ferrari & Clinkenbeard, 1999, p.6; Sternberg, Ferrari, Clinkenbeard & Grigorenko, 1996, p.133) os índices KR-20 obtidos foram os seguintes:

- I.Analítica .63, I. Prática .48 e I. Criativa .62

- Numa outra amostra (N=264 estudantes pré-universitários e universitários recém admitidos a uma Universidade privada de Lima, Perú, média de idades 17 anos) (Carrasco, 2000, p.18), foi aplicada uma versão do *STAT (H)* (1993) traduzida para espanhol (12 itens para cada variável global) e foram obtidos os seguintes coeficientes:
  - I.Analítica .50, I. Prática .35 e I. Criativa .51
  - Áreas Verbal .21, Quantitativa .64 e Figurativa .46 [24]
- Num estudo mais recente, efectuado com o *STAT-R (H)* (2002) numa amostra de estudantes do primeiro ano universitário ou do último ano secundário (N=990) (Sternberg & The Rainbow Project Collaborators, 2006, p.332), foram obtidos os seguintes resultados:
  - coeficientes Alfa de Cronbach: I.Analítica .67, I. Prática .56 e I. Criativa .72 [25]
  - coeficientes de precisão dos sujeitos (Rasch): I.Analítica ..59, I. Prática .53 e I. Criativa .60.

Os índices persistentemente baixos de consistência interna, mesmo aqueles que sendo relativos aos três domínios – analítico, prático e criativo – ou às três áreas de conteúdo – verbal, quantitativa e figurativa – envolvem mais itens [12 na versão do *STAT (H)* (1993), 15 na versão *STAT-R (H)* (2002)] – são interpretados por Sternberg e colaboradores como reflexo da natureza compósitas destes resultados, os quais em cada domínio de processamento incluem três áreas de conteúdo, e em cada área de conteúdo incluem três domínios de processamento (Sternberg, Grigorenko, Ferrari & Clinkenbeard, 1999). Embora este argumento encontre apoio nas posições menos ortodoxas de alguns autores, que admitem que índices menores de consistência interna são aceitáveis quando o construto sob avaliação é conceptualmente heterogéneo (Messick, 2000; Raven, 1989), a confirmarem-se estes níveis de consistência, eventualmente aliados a outras limitações metrológicas detectáveis na

[24] Consistência interna do resultado global de todo o teste (36 itens) .67.
[25] Consistência interna do resultado global de todo o teste (45 itens) .79.

aplicação do modelo de Rasch, colocam-se desde logo sérios entraves à investigação da inteligência com recurso ao *STAT.*

- Quanto aos dados de **VALIDAÇÃO**, Sternberg apresenta evidências de validação intra-conceito (que designa de "validade interna") e evidências de validação inter-conceito (que designa de "validade externa") (Dickes, Tournois, Flieller & Kop, 1994).
  - Validação intra-conceito (estudos de análise factorial exploratória e confirmatória):
    - No único estudo em que o método de análise factorial exploratória aparece tratado na bibliografia sobre a validação do *STAT (H)* (Sternberg, Ferrari, Clinkenbeard & Grigorenko, 1996, p.134) foi aplicado o método de componentes principais ao nível dos itens, com rotação varimax: foram extraídos nove factores com valores próprios iguais ou superiores a .98, cada teste saturando num factor em torno de .9 e apresentando saturações em torno de .2 nos restantes factores. Estes resultados foram interpretados como confirmação de que o *STAT (H)* não é mais uma medida de *g*, uma vez que cada subteste parece envolver uma combinação única processo-conteúdo.
    - No estudo de análise factorial confirmatória citado a propósito da descrição das partes constituintes do *STAT (H),* concluiu-se que o modelo mais ajustado à estrutura dos dados era o modelo baseado na Teoria Triárquica, isto é, que postulava nove factores de primeira ordem e três factores de segunda ordem (analítico, prático e criativo), ainda que seja admitido que outros modelos alternativos apresentaram índices de ajustamento muito próximos (Sternberg, Catejón, Prieto, Hautamäki & Grigorenko, 2001, p.13; Sternberg, The Rainbow Project Collaborators et al., 2004, p189). O ajustamento deste modelo foi melhorado quando se permitiu a correlação das variáveis de erro. Os índices de ajustamento deste último modelo, obtidos em amostras conjuntas (segundo os autores, as amostras foram reunidas duas a duas dado que, por limitações de memória do programa, não foi possível tratar, como se pretendia, as três amostras reunidas):
      - χ^2/gl=2.75, GFI=.91, RMSEA=.033, CFI=.83 e RFI=.71 (Finlândia+USA);

- χ^2/gl=2.21, GFI=.99, RMSEA=.027, CFI=.86 e RFI=.74 (Finlândia+Espanha)
- χ^2/gl=0.52, GFI=1.00, RMSEA=.000, CFI=1.0 e RFI=.57 (USA+Espanha).

- No estudo de análise factorial confirmatória conduzido apenas com a subamostra espanhola do estudo anterior (Sternberg, Prieto & Castejón, 2000) (N=240, estudantes do ensino secundário), conclui-se que embora o modelo baseado na Teoria Triárquica seja o mais ajustado aos dados (χ^2/gl=1.30, CFI=.806) o ajustamento está longe de ser perfeito, sendo possível melhorar o ajustamento ao permitir as correlações dos erros, o que os autores não consideram, contudo, tecnicamente recomendável (p.646).
- Num outro estudo do *STAT (H)* (1993) em que foi aplicado o método de análise factorial confirmatória (Carrasco, 2000) conclui-se favoravelmente em relação ao ajustamento da estrutura dos dados ao modelo teórico – Analítica, Prática e Criativa –, mas reconhece-se que os dados se ajustam igualmente à estrutura baseada nas áreas de conteúdo – Verbal, Quantitativa e Figurativa:
  - Modelo Triárquico (A,P,C): χ^2/gl=.978, GFI=.98, RMSEA=.000, AGFI=.96
  - Modelo de Áreas de Conteúdo (V,Q,F): χ^2/gl=.949, GFI=.98, RMSEA=.000, AGFI=.96
- Finalmente, no estudo exploratório da estrutura factorial do *STAT-R (H)* (2002), em conjunto com novas medidas baseadas na teoria triárquica (a apresentar no CAPÍTULO 9) (N=990, universitários e pré-universitários), verificou-se que as três medidas, Analítica, Prática e Criativa, apresentaram correlações elevadas entre si
  - I.Analítica – I.Prática : .62
  - I.Analítica – I.Criativa : .57
  - I.Prática – I.Criativa : .61

  e definiram um factor (saturações respectivamente de .80, .81 e .73) separado dos factores definidos pelas restantes medidas, o que é justificado pelos autores como consequência da metodologia comum

(resposta por escolha múltipla) e dos conteúdos comuns aos três domínios de processamento (verbal, quantitativo e figurativo); este factor é, então, lido como um "factor comum metodológico" que suplanta os contributos analítico, prático e criativo das medidas do *STAT* (Sternberg & The Rainbow Project Collaborators, 2006, p.340). Na análise confirmatória deste estudo, o modelo mais ajustado foi o que incluía como variável latente o *STAT* (conjunto das três medidas) "sobretudo encarado como uma medida de *g* (i.e., analítica)" (p.340).

- Validação inter-conceito (validade convergente e discriminante):
  - Correlações entre o *STAT (H)* (1993) e outros testes (Sternberg, 1997d, p.22; Sternberg & Clinkenbeard, 1995, p.257; Sternberg, Ferrari, Clinkenbeard & Grigorenko, 1996, p. 132):
    - *Watson-Glaser Critical Thinking Appraisal*. Correlações: com a I. Analítica .50, com a I.Prática .32 e com a I. Criativa .53;
    - *Concept Mastery Test*: Correlações: com a I. Analítica .49, com a I.Prática .21 (não significativa) e com a I. Criativa .43;
    - *Cattell Culture-Fair test of g*: Correlações: com a I. Analítica .50, com a I.Prática .36 e com a I. Criativa .55;
    - *Teste de insight-criativo*, elaborado por Sternberg e colaboradores: Correlações: com a I. Analítica .47, com a I.Prática .21 e com a I. Criativa .59.
  - Correlações entre *STAT-R (H)* (2002) e outros testes (Sternberg & The Rainbow Project Colaborators, 2006):
    - *Scholastic Assessment Test – Mathematic* (*SAT-M*): Correlações - com I. Analítica .62, com I.Prática .57 e com I.Criativa .60;
    - *Scholastic Assessment Test – Verbal* (*SAT-V*): Correlações - I. Analítica .53, com I.Prática .53 e com I.Criativa .55.
  - Estudos educacionais:
    - O *STAT (H)* (1993) foi o instrumento de avaliação aplicado para recrutamento dos participantes no estudo em que foi aplicada uma metodologia *ATI* – estudo das interacções da inteligência funcional com os métodos de ensino, dirigidos às

> três formas de inteligência ou à memorização. Tratou-se de um estudo em que foi ministrado um curso de Verão de Introdução à Psicologia a estudantes do ensino secundário recrutados maioritariamente nos Estados Unidos, no qual os mesmos conteúdos foram leccionadas com recurso a metodologias concebidas com base nos três domínios – analítico, prático e criativo – e ainda na memorização, Os resultados apresentados na literatura (Sternberg & Clinkenbeard, 1995; Sternberg, Ferrari, Clinkenbeard & Grigorenko, 1996; Sternberg, Grigorenko, Ferrari & Clinkenbeard, 1999) demonstraram que o teste se mostrou útil para a identificação das potencialidades dos estudantes no quadro da Teoria Triárquica (alguns resultados metrológicos obtidos neste estudo foram já abordados atrás).

Em suma, algumas evidências provenientes dos estudos de validação tendem a apoiar a validade de construto das medidas proporcionadas pelo *STAT (H)* (1993), mas estão longe de dar do instrumento uma leitura conceptual muito sólida. Não só os índices de precisão se mostram frágeis, como a estrutura das variáveis, embora tenda a ajustar-se ao modelo teórico esperado, também se ajusta, com grau de eficácia semelhante, a outros modelos conceptuais. Na sua versão mais recente, o *STAT-R (H)* (2002) começa a ser até visto no conjunto como uma medida de inteligência analítica ou mesmo de *g*. A fragilidade das interpretações e das conclusões de Sternberg e colaboradores a propósito da validação do *STAT (H)* e da própria Teoria Triárquica valeu, aliás, críticas veementes e um aceso debate na literatura [ver número especial da revista *Intelligence*, em 2003, *31* (4)[26]]. A revisão que deu origem ao *STAT-R (H)* (2002) terá talvez procurado ultrapassar algumas destas limitações, mas não foi dada suficiente atenção à replicação dos estudos metrológicos, tendo o autor enveredado mais pela exploração de outras formas de avaliação da inteligência funcional (ver CAPÍTULO 9).

Uma vez reconhecidas as limitações do *STAT (H)* do ponto de vista metrológico, sobrevém a questão: porquê escolher este teste como representante do paradigma sistémico de avaliação da inteligência na presente investigação? A resposta é simples: porque é a única técnica disponível que proporciona a medição da inteligência no quadro de um modelo sistémico (logo, a única passível de fornecer dados adequados à articulação com a *WAIS-III*). Com efeito, ao retomar os modelos sistémicos alternativos, facilmente se compreende esta opção: para Gardner a avaliação da inteligência não pode basear-se na abordagem psicométrica, relativamente à qual é convictamente crítico, porque,

[26] Este debate será tratado a propósito da discussão dos resultados do presente estudo, no CAPÍTULO 9.

enquanto conceito contextual, a inteligência só pode ser avaliada em situações "reais", através de tarefas comuns de confronto com situações e problemas colocados em contexto (Gardner, 1993, 1999, 2006). Por outro lado, embora não assumindo uma posição tão explicitamente contrária à medida diferencial, Ceci não operacionalizou a sua teoria em nenhum instrumento de medição da inteligência – tal opção surge francamente justificada pelos conceitos que a teoria inclui e pela orientação desenvolvimentista que adopta. Por fim, a teoria da Inteligência Emocional, ainda que proporcionando técnicas de medida baseadas em sólido desenvolvimento metrológico, não constitui nem pretende constituir um modelo com carácter integrativo, como assinalado anteriormente, mais se inscreve numa concepção ecléctica de investigação e de avaliação da inteligência (ver CAPÍTULO 2, pp.127-128) e, por isso mesmo, os instrumentos que fornece não são os mais adequados à prossecução dos objectivos da presente investigação. Daí que a opção tenha sido a de procurar aperfeiçoar o teste, na sua versão portuguesa, esperando que as modificações introduzidas aquando da sua revisão – *STAT-R (H)* (2002) – pudessem ter contribuído para aumentar a sua robustez metrológica, e fazendo desse trabalho de estudo experimental e sucessivo aperfeiçoamento parte integrante da presente investigação.

4.2.2. Amostragem de Indivíduos

Nenhuma das amostras utilizadas na presente investigação pretende ser tomada como representativa da população portuguesa adulta, pelo que não seguiu exaustivamente os procedimentos recomendados na literatura (Miranda, 1983) e aplicados nos grandes trabalhos de aferição nacional de técnicas diferenciais (como, por exemplo, Marques, 1969; Miranda, 1982; Pinto, 2002; Seabra-Santos, 1998; Wechsler, 2003) nem os procedimentos rigorosos de uma amostragem ao acaso. Mesmo admitindo que as possibilidades de generalização das conclusões ficam seriamente limitadas por esta opção, que implica enorme prudência na interpretação e discussão dos resultados, imperativos de ordem técnica e de ordem prática determinaram que fosse a mais viável.

De um ponto de vista técnico, o investimento que a organização de amostras representativas de populações nacionais exige, em termos de recursos humanos e materiais, justifica-se quando está em causa o estudo de uma população numa variável ou num conjunto de variáveis, muito em especial quando se pretende construir normas para posterior obtenção de resultados individuais reportados aos parâmetros populacionais. Ora, este momento do desenvolvimento de uma investigação pressupõe o estudo prévio e aperfeiçoamento dos instrumentos de medida dessas variáveis. É precisamente nesta fase de estudo prévio, quer da *WAIS-III*, quer do *STAT-R (H)*, que esta investigação se inscreve, posto que ambas as técnicas foram traduzidas e adaptadas no âmbito deste projecto de investigação, aconselhando a prudência científica que, apesar das sucessivas tentativas de aperfeiçoamento, elas sejam por ora consideradas em fase de estudo experimental. Por outro lado, o propósito desta

investigação situa-se mais num quadro conceptual do que metodológico, ou seja, pretende-se questionar como tende a organizar-se um conjunto de medidas, testando a proximidade dessa organização a determinados modelos teóricos, mais do que estudar como tendem essas medidas a distribuir-se na população; da utilização de amostras não representativas da população decorre apenas que há que restringir o âmbito das conclusões em função das características das amostras, que não devem nunca ser perdidas de vista na discussão dos resultados, e há que evitar generalizar para a população as estatísticas obtidas nas amostras. Por fim, importa assinalar que uma das metodologias adoptada – a análise de itens de acordo com modelos de traço latente (adiante caracterizados) – oferece a enorme vantagem de proporcionar parâmetros dos itens independentes das características das amostras, o que legitima, nesta fase experimental, a utilização de amostras não rigorosamente representativas da população.

De um ponto de vista prático, o facto de estarem sob estudo não uma mas duas técnicas diferenciais, ambas de aplicação longa (no mínimo uma hora e meia cada uma) e uma delas de aplicação individual (a duração da aplicação depende do ritmo do examinado e ocupa em exclusivo um examinador), implicaria enorme investimento em recursos humanos, caso se pretendesse estender as aplicações a todo o país. Acresce que no Estudo Principal seria extremamente difícil o recrutamento de participantes que se voluntariassem, a nível nacional, para cerca de três horas de resolução de técnicas diferenciais (para responder aos dois testes).

Assim, as amostras observadas no presente estudo foram recolhidas informalmente, a partir da delimitação prévia de determinados requisitos básicos – domínio da língua e da cultura portuguesas, limites etários, ambos os sexos, diversidade de níveis educacionais e sócio-culturais – apenas se tendo procurado, ao longo de cada estudo, ir corrigindo assimetrias que pudessem contribuir para desequilibrar excessivamente as amostras: por exemplo, na amostra do Estudo Principal, em certa fase do desenvolvimento experimental houve que intensificar o recrutamento de participantes do sexo masculino para procurar corrigir o desequilíbrio que espontaneamente estava a delinear-se entre as subamostras feminina e masculina; ou, nas amostras dos Ensaios Experimentais, incentivou-se os examinadores a procederem à recolha das amostras fora da zona da "Grande Lisboa", ou à recolha de participantes não estudantes e não licenciados, o que apenas em parte foi conseguido. Apesar destes cuidados, a caracterização das amostras, que será efectuada aquando da apresentação dos resultados dos diferentes estudos, denuncia enviesamentos em relação à população nacional, designadamente no nível de escolaridade, no nível sócio-profissional (mais elevado nas amostras do que a nível nacional) e na distribuição geográfica da população (sobretudo concentrada na região de Lisboa e Vale do Tejo e na área litoral).

O recrutamento dos participantes contou, regra geral, com a colaboração dos alunos da disciplina de Psicologia Diferencial[27] (Licenciatura em Psicologia, Faculdade de Psicologia e de Ciências da Educação, Universidade de Lisboa) e teve lugar no âmbito da realização de trabalho prático facultativo de colaboração em projectos de investigação diferencial. Em cada estudo foram definidas directrizes gerais, entregues a cada colaborador na forma de documento escrito (ANEXO 1) em aula prática expressamente dedicada às questões metodológicas envolvidas no estudo em causa. Entre outros aspectos, estas directrizes continham indicações estritas sobre os critérios a que deveria obedecer o recrutamento dos participantes, os quais contemplavam, as seguintes variáveis:

- Sexo: participantes de ambos os sexos, se possível em número igual;
- Idade: 16 anos e 0 meses a 80 anos e 0 meses; diversificação etária, dentro do possível. As amostras foram categorizadas em sete níveis etários[28]:
  - 1: ≤ 19 anos;
  - 2: 20 – 29 anos;
  - 3: 30 – 39 anos;
  - 4: 40 – 49 anos;
  - 5: 50 – 59 anos;
  - 6: 60 – 69 anos;
  - 7: 70 – 79 anos[29].
- Nível de Escolaridade: diversificação do nível de escolaridade, procurando que pelo menos metade dos participantes angariados para as amostras tivesse 12 ou menos anos de escolaridade (exigência básica: domínio de competências de leitura e escrita). A categorização dos níveis de escolaridade seguiu a seguinte classificação:
  - 1: < 4 anos de escolaridade (1º ciclo ensino básico incompleto);
  - 2: 4 – 5 anos de escolaridade (1º ciclo ensino básico a 2º ciclo incompleto);
  - 3: 6 – 8 anos de escolaridade (2º ciclo ensino básico a 3º ciclo incompleto);
  - 4: 9 – 11 anos de escolaridade (3ºciclo do ensino básico ou escolaridade obrigatória a estudos secundários incompletos);

[27] Aproveita-se para expressar o devido reconhecimento aos alunos que frequentaram a disciplina de Psicologia Diferencial nos anos lectivos de 2002/2003, 2003/2004 e 2004/2005, pelo seu precioso contributo para a angariação de participantes e a recolha de dados dos ensaios experimentais da *WAIS-III* e do *STAT-R (H).*

[28] Esclareça-se que as categorizações que a seguir se identificam, em relação às diversas variáveis demográficas contempladas, apenas serviram para descrever e caracterizar as amostras; como referido, não constituíram critérios de estratificação das amostras.

[29] Não se registou em nenhum dos estudos qualquer participante com 80 anos.

- 5: 12 – 14 anos de escolaridade (12º ano ou estudos secundários completos a frequência de ensino superior)
- 6: 15 – 16 anos de escolaridade (curso médio/ bacharelato/curso politécnico completos);
- 7: ≥ 17 anos de escolaridade (licenciatura a estudos pós-graduados).

- Profissão: dentro do possível, diversificação das áreas de actividade dos participantes entre os sectores de actividade económica primário, secundário e terciário. Procurou-se algum controlo do número de estudantes nas amostras (os participantes mais fáceis de recrutar mas pouco representativos da população, sobretudo para a investigação da cognição humana), exigindo que esse número nunca ultrapassasse metade dos participantes angariados por cada colaborador. A categorização das amostras quanto à profissão inspirou-se na Classificação Nacional das Profissões (Versão de 1994) (Instituto do Emprego e Formação Profissional, 1994) e assumiu as seguintes categorias:
  - 0: Estudantes;
  - 1: Quadros superiores da administração pública, dirigentes e quadros superiores das empresas;
  - 2: Especialistas das profissões intelectuais e científicas;
  - 3: Técnicos e profissionais de nível intermédio;
  - 4: Pessoal administrativo e similares;
  - 5: Pessoal dos serviços e vendedores;
  - 6: Agricultores e trabalhadores qualificados da agricultura e pescas;
  - 7: Operários, artífices e trabalhadores similares
  - 8: Operadores de instalações e máquinas e trabalhadores da montagem
  - 9: Trabalhadores não qualificados
  - 10: Militares;
  - 11:Indefinidos: trabalhadores não classificáveis noutros grupos (domésticas; reformados/ funcionários públicos/ desempregados sem especificação de área de actividade).
- Região/Área/Tipo de residência: procurou-se estimular a angariação de participantes fora da zona da "Grande Lisboa", por exemplo, sugerindo o recrutamento de participantes nas regiões/áreas de residência originárias dos colaboradores, sempre que estes se encontravam deslocados a estudar em Lisboa. Para o efeito, procurou-se

estimular o aproveitamento de períodos de férias escolares, sobretudo para os estudantes deslocados em Lisboa, para a implementação do recrutamento e avaliação dos participantes. A classificação do local de residência adoptou as seguintes categorizações[30]:

Região de residência	Área de residência	Tipo de residência
1: Norte	1: Litoral	1: Grande Centro Urbano
2: Centro	2: Interior	2: Concelho Urbano
3: Lisboa e Vale do Tejo	3: Ilhas	3: Outros Concelhos
4: Alentejo		
5: Algarve		
6: Açores		
7: Madeira		

As aplicações do Estudo Principal decorreram, na maioria, nas instalações da Faculdade de Psicologia e de Ciência da Educação (Alameda da Universidade, Lisboa), o que significou que os participantes se deslocavam propositadamente ao local de aplicação, tendo prévio conhecimento de que lhes seria pedido que dedicassem uma manhã ou uma tarde (cerca de três horas e meia) à realização de provas psicológicas. Esta circunstância terá sem dúvida contribuído para o carácter seleccionado da amostra, que possivelmente assumiu características muito particulares, do ponto de vista da motivação como do nível educacional e cultural, as quais é necessário serem devidamente atendidas na análise e discussão dos resultados desse estudo.

4.2.3. Organização dos Planos Experimentais

O planeamento e a organização da experiência são acima de tudo orientados pelo problema e pelas hipóteses experimentais, que derivam directamente do problema. Uma vez identificadas as variáveis e definidos os parâmetros da selecção dos indivíduos, torna-se agora possível traduzir o problema em hipóteses experimentais, relativas ao Estudo Principal da presente investigação, para de seguida caracterizar em linhas gerais os planos experimentais e as técnicas de tratamento de dados adoptadas.

[30] Foram seguidos os critérios mais recentemente adoptados na organização de amostras portuguesas para aferição de testes (Pinto, 2002; Wechsler, 2003) os quais se fundamentam em conceitos estabelecidos pelo Instituto Nacional de Estatística (www.ine.pt).

1) **HIPÓTESES EXPERIMENTAIS**

Tomando como ponto de partida o problema delineado, e considerando as medidas proporcionadas pelas duas técnicas diferenciais sob estudo, é possível derivar da teoria as seguintes hipóteses relativas à estrutura das variáveis, tal como são medidas pelos dois instrumentos de avaliação sob estudo, a *WAIS-III* versão experimental portuguesa (2002) e o *STAT-R(H)* versão experimental portuguesa (2004) :

HIPÓTESE 1: a inteligência geral é medida por ambos os instrumentos e representa uma dimensão única em que saturam todas as variáveis, a saber, todos os subtestes da *WAIS-III aplicados* e todas as partes do *STAT-R(H)*, ou seja, todas as formas de inteligência funcional ou domínios de processamento mental identificados na Teoria Triáqruica (analítica, criativa e prática).

- Esta hipótese corresponde à perspectiva subscrita por muitos autores do passado, desde Spearman, e por nomes notáveis da actualidade (ver Nyborg, 2003) que subscrevem a noção de *g* como um dos grandes triunfos da ciência psicológica.
- Representa também a concepção de inteligência de Wechsler, posto que a identificação de uma única dimensão subjacente à variabilidade de resultados nos dois testes confirmaria a presença de uma única forma de inteligência, que se manifesta sempre, independentemente dos conteúdos ou das situações problema em que é avaliada, o que conduziria à confirmação do valor dos instrumentos compósitos de avaliação da inteligência geral, como a *WAIS-III.*
- De acordo com Wechsler (1975), "o que medimos com os testes não coincide com o que os testes medem – não é informação, nem percepção espacial, nem capacidade de raciocínio. O que os testes de inteligência medem, o que esperamos que meçam, é algo muito mais importante: a capacidade do indivíduo para compreender o mundo à sua volta e os seus recursos para lidar com os seus desafios".
- O que esta hipótese prevê é que os novos tipos de testes concebidos por Sternberg nada vêm acrescentar à avaliação da inteligência assim concebida, posto que mais não serão do que outras tantas formas de medir uma mesma inteligência global.

HIPÓTESE 2: os testes de Inteligência Analítica do *STAT-R (H)* saturam em conjunto com os subtestes da *WAIS-III* num factor distinto dos factores em que saturam a Inteligência Criativa e a Inteligência Prática.

- Esta hipótese deriva directamente da Teoria Triárquica da Inteligência de Sternberg e corresponde a admitir que o conceito tradicional de inteligência geral apenas é "geral" porque a gama de funcionamento envolvida nos testes de inteligência clássicos, como a *WAIS-III*, é estreita e demasiado homogénea, por corresponder ao espectro do funcionamento cognitivo recompensado e treinado no âmbito das actividades escolares, aquele em que a escola promove o "desenvolvimento de mestria". Se essa gama for ampliada, de acordo com a Teoria Triárquica, *g* deixará de emergir.
- Sternberg sublinha: "[…] apresentar bons resultados nos três aspectos da inteligência não corresponde a um nível elevado de "g" psicométrico, ou de aptidão geral do tipo da divulgada como base da inteligência por muitos psicólogos desde Spearman (1927) até Jensen (1972) ou Herrnstein e Murray (1994). Os testes de inteligência do tipo dos utilizados por estes investigadores são encarados na Teoria Triárquica como medindo alguns aspectos da inteligência analítica, dificilmente tocando sequer as inteligências criativa e prática." (Sternberg, Ferrari, Clinkenbeard & Grigorenko, 1996, pp.130-131). A Hipótese 2 é inspirada precisamente por esta convicção.
- A confirmação desta hipótese sugeriria não necessariamente o abandono dos instrumentos clássicos, mas o reconhecimento do reducionismo a que conduz a sua utilização exclusiva e a necessidade de ampliação da medida da inteligência pela avaliação de outras áreas de funcionamento (criativa e prática).

HIPÓTESE 3: os testes de conteúdo Verbal e Quantitativo do *STAT* saturam no factor de Compreensão Verbal da *WAIS-III,* enquanto os testes de conteúdo Figurativo saturam no factor de Organização Perceptiva.

- Esta hipótese equivale ao afirmar de um modelo de inspiração multifactorial, em que as medidas se organizam em função das aptidões para lidar com determinadas formas de representação mental da

informação ou conteúdo – simbólico (palavras ou números) ou perceptivo (figurativo).

- Esta hipótese corresponde a por em causa a noção de Sternberg de que os processos envolvidos na resolução de problemas são de natureza mais universal (comuns a diversas culturas), e mais transversal (comuns a várias tarefas), do que os conteúdos sobre os quais operam. De acordo com o autor, "o que difere entre as áreas de conteúdo, segundo a Teoria Triárquica, não é o conjunto de processos de tratamento da informação envolvidos, mas os conteúdos mentais ou representações que são utilizadas nas diferentes áreas. Por consequência, teremos de assumir que a aplicação destes três aspectos da inteligência – analítico, prático e criativo – tem um alcance muito mais psicologicamente fundamental do que o modo da representação dos conteúdos a que os vários aspectos da inteligência se aplicam – verbal, quantitativo ou figurativo." (Sternberg, Castejón, Prieto, Hautamäki & Grigorenko, 2001, p.2).
- Sublinhe-se que nesta hipótese não se supõe ver emergir um factor geral, mas antes factores de grupo – pelo menos dois, um simbólico e outro perceptivo, eventualmente acrescidos de um factor de memória, e/ou de velocidade perceptiva, e/ou burocrático-motor, estes definidos a partir de alguns subtestes da *WAIS-III* – numa réplica da estrutura identificada em outros estudos factoriais das aptidões (Pinto, 2002).

HIPÓTESE 4: a estrutura das medidas dos dois testes emerge como uma hierarquia em que tanto se identifica g, explicando uma parte substancial da variância dos resultados, como emergem factores de grupo. Esta hipótese pode subdividir-se em duas:

HIPÓTESE 4.1: para além de *g*, identificam-se factores de grupo coincidentes com os aspectos da inteligência postulados pela Teoria Triárquica ou pela Teoria da Inteligência Funcional de Sternberg.

- Embora pondo em causa a crítica de Sternberg à gama limitada de funcionamento avaliada pelos testes convencionais de inteligência, uma vez que *g* abrangeria afinal todas as formas de inteligência sugeridas pelo autor, esta hipótese constitui uma versão modificada da Hipótese 2 e, a confirmar-se, apoiaria parcialmente a Teoria Triárquica da Inteligência,

designadamente ao relevar as três formas de inteligência em detrimento dos três tipos de conteúdos.

- A confirmação desta hipótese constituiria também fundamento para o apuramento de um resultado global no *STAT-R (H),* mas poria em causa algumas das afirmações de Sternberg, designadamente, a de que o equilíbrio entre diversas formas de inteligência é mais determinante para a definição da inteligência do que a manifestação de um nível geral de funcionamento. Ainda assim, poderia ser argumentado que a instância de nível superior e de carácter geral correspondesse a funções executivas encarregues da gestão dos recursos mentais consubstanciados nas três formas de inteligência funcional.

HIPÓTESE 4.2: para além de g, identificam-se factores de grupo coincidentes com os tipos de conteúdos ou formas de representação mental da informação – Verbal, Quantitativa e Figurativa – presentes no *STAT* e também na *WAIS*.

- Esta hipótese, que corresponde a uma estrutura relativamente próxima dos modelos hierárquicos das aptidões, poria seriamente em questão a Teoria Triárquica, quer pela emergência de *g*, quer pela menor pertinência das três formas de inteligência postuladas por Sternberg. Constitui uma versão modificada da Hipótese 3 e, a confirmar-se, daria apoio à perspectiva clássica da inteligência – cuja mais recente e consensual representação se encontra na Teoria dos Três Estratos de Carroll – na qual a inteligência funcional consiste meramente numa forma diferente de organizar a variância dos resultados obtidos na aplicação de medidas da cognição humana.

2) **PLANOS EXPERIMENTAIS**

Tendo em vista o teste das hipóteses experimentais formuladas, a organização do projecto de investigação comportou duas etapas: a primeira, Ensaios Experimentais, dedicada à tradução, adaptação e aperfeiçoamento das versões portuguesas experimentais das técnicas diferenciais escolhidas para representar os dois paradigmas sob estudo, a *WAIS-III* e o *STAT-R (H)*; a segunda, Estudo Principal, dedicada ao problema delimitado no início deste capítulo, e ao teste das hipóteses experimentais que acabam de se estabelecer. A organização dos planos experimentais comportou, por consequência, diversas fases e diferentes estudos, cujos objectivos e natureza geral se caracterizam

de seguida, de acordo com um critério cronológico, sendo os procedimentos específicos adoptados na sua implementação descritos em maior detalhe no CAPÍTULO 5.

1. ADAPTAÇÃO E ESTUDOS EXPERIMENTAIS DA *WAIS-III* E DO *STAT-R (H)*

Nesta etapa preparou-se os instrumentos para representarem os paradigmas diferencial e sistémico de avaliação da inteligência humana, no quadro dos objectivos delineados para o presente projecto. Este trabalho comportou as seguintes fases:

- ***WAIS-III*** : **ESCALA DE INTELIGÊNCIA DE WECHSLER PARA ADULTOS – 3ª EDIÇÃO: VERSÃO EXPERIMENTAL PORTUGUESA (2002)**
  - 2000 – 2002:
    - Tradução: preparação do Manual Experimental 1 (Wechsler, 2002a);
    - Revisão da tradução: preparação do Manual Experimental 2 (Wechsler, 2002b);.
  - 2001/2002:
    - Organização do Ensaio Experimental;
    - Ensaio Experimental (N=225) (aplicações entre Março e Junho/2002);
    - Estudo dos Critérios de Classificação (n=76).
  - 2003/2004:
    - Preparação de aditamento ao Manual Experimental 2: "Critérios de Classificação dos Subtestes de Vocabulário, Semelhanças e Compreensão" (2004)
    - Estudo do acordo inter-avaliadores e da estabilidade temporal (aplicações entre Março e Junho de 2004);
      - Acordo inter-avaliadores (N=100);
      - Estabilidade temporal (n=86);
      - Consistência Interna (N=100)
    - Ensaio Experimental: Cotação, registo de dados e análise de resultados:
      - Teoria Clássica dos Testes:
        - Análise de itens;
        - Estudo da consistência interna.
      - Análise Factorial Exploratória.
      - Análise no quadro dos Modelos de Traço Latente
      - Análise Factorial Confirmatória.

- Revisão do Manual Experimental 2.

- ***STAT-R (H):* TESTE TRIÁRQUICO DE APTIDÕES DE STERNBERG - REVISTO (NÍVEL H) VERSÃO EXPERIMENTAL PORTUGUESA (2002, 2003, 2004)**
  - 2002/2003:
    - Tradução: preparação das Instruções de Aplicação, do Caderno de Teste e da Folha de Respostas (Afonso, 2002,b,c);
    - Estudo Piloto (N=66) (aplicações entre Dezembro/2002 e Janeiro/2003);
    - Registo de dados, cotação automática e análise de resultados;
    - Revisão da Tradução: preparação de nova versão das Instruções de Aplicação, do Caderno de Teste e da Folha de Respostas (Afonso, 2003c,d);
    - 1º Ensaio Experimental (N=820) (aplicações entre Março e Junho/2003).
  - 2003/2004:
    - Registo de dados, cotação automática e análise de resultados:
      - Amostra total (N=820);
      - Amostra com escolaridade igual ou superior ao 12º ano (n=406);
    - Revisão da Tradução: preparação de nova versão do Caderno de Teste e da Folha de Respostas (Afonso, 2004b,c);
    - 2º Ensaio Experimental (N=370) (aplicações entre Março e Junho/2004);
    - Registo de dados, cotação automática e análise de resultados.
  - 2004/2005:
    - Ensaio sem tempo limite (N=487) (aplicações entre Março e Junho/2005);
    - Registo de dados, cotação automática.
  - 2005/2006:
    - Ensaio sem tempo limite: análise de resultados.

2. ESTUDO PRINCIPAL: *WAIS-III + STAT-R (H)*

Nesta etapa, os dois instrumentos foram utilizados em conjunto na mesma amostra, tendo em vista o estudo das correlações entre as medidas (estratégia correlacional) e a aplicação de métodos

multivariados, designadamente análise factorial confirmatória para teste do ajustamento da estrutura das medidas a diferentes concepções teóricas da inteligência. Esta parte do trabalho organizou-se nas seguintes fases:

- 2004/2005:
  - Organização da campanha de aplicações do Estudo Principal: recrutamento e formação dos examinadores; planeamento e preparação de materiais e dos locais de aplicação; recrutamento dos participantes.
  - Estudo Principal (N=250)
    - Aplicações (Novembro/2004 – Fevereiro/2005)
    - Cotação e registo de dados
- 2005/2006:
  - Estudo Principal: Análise de Resultados
    - Novo estudo metrológico dos instrumentos:
      - Teoria Clássica dos Testes:
        - Análise de itens;
        - Estudo da consistência interna.
      - Análise Factorial Exploratória.
      - Análise no quadro dos Modelos de Traço Latente.
      - Análise Factorial Confirmatória.
    - Estudo da Estrutura das variáveis:
      - Análise Factorial Exploratória;
      - Análises de Unidimensionalidade (Modelos de traço latente);
      - Análise Factorial Confirmatória.

Na lista precedente, encontram-se sublinhados os estudos efectuados no âmbito da presente investigação: os Ensaios Experimentais realizados separadamente com cada uma das provas, ao terem por finalidade ensaiá-las na população portuguesa e estudá-las do ponto de vista metrológico, adoptaram planos de experiência comuns neste tipo de investigação: as técnicas foram administradas cumprindo rigorosamente as condições de aplicação estipuladas no Manual (no caso da *WAIS-III*) ou nas Instruções de Aplicação (no caso do *STAT-R (H)*), apenas se introduzindo algumas alterações quando se mostraram pertinentes face aos propósitos de cada estudo. Por exemplo, no Ensaio Experimental da *WAIS-III*, optou-se pela aplicação de todos os subtestes (catorze), tendo em vista estudá-los na população portuguesa e escolher posteriormente os que, sendo metrologicamente mais

robustos, se mostravam mais adequados para utilização no Estudo Principal; para evitar que as aplicações se tornassem demasiado longas, neste estudo não foram aplicados os itens de inversão de maneira sistemática, apenas quando se impunha devido à ocorrência de insucessos nos primeiros itens aplicados, nem foi implementada a prática, comum nas fases preliminares de estudo das técnicas diferenciais, de ampliar o critério de paragem da aplicação em cada subteste (prevenindo que os itens na versão traduzida não tenham mantido a ordem de dificuldade crescente que tinham na versão original, o que prejudica os examinados). Assim, no Ensaio Experimental da *WAIS-III* a informação sobre os primeiros itens, os "itens de inversão", não foi suficiente para os incluir na análise de itens e nos estudos de precisão; e a informação sobre as percentagens de acertos também não foi suficiente para permitir alterar com segurança a sua ordenação. Este tipo de análises foi implementado mais tarde, por ocasião do estudo de acordo inter-avaliadores e de estabilidade temporal, em que todos os itens de inversão foram sistematicamente aplicados e em que se ampliou o critério de paragem, pois neste estudo não foram já aplicados catorze subtestes mas apenas onze.

Uma modificação importante das condições de aplicação estandardizada foi introduzida num dos estudos do *STAT-R (H),* o estudo "sem tempo limite". Como se verá adiante, nalgumas partes da prova, o tempo limite máximo de 8 minutos, já incluindo prolongamento, não se mostra suficiente para que uma percentagem elevada de examinados terminem a tarefa; decorre que se dispõe sempre de muito menos informação para análise de itens relativamente aos últimos, do que aos primeiros (uma percentagem muito elevada dos últimos itens nem chega a ser tentada pelos examinados, de onde o insucesso não significa que sejam difíceis, mas tão só que não foram atingidos). A aplicação sem tempo limite procurou obviar a esta dificuldade e ao mesmo tempo preparar os dados para poderem ser tratados no quadro dos modelos de traço latente[31].

No Estudo Principal, as aplicações foram planeadas de forma a alternar a ordem das duas provas – aproximadamente metade da amostra respondeu primeiro à *WAIS-III* (56,4%) e a outra metade começou pelo *STAT-R (H)* (43,6%) – e seguiram as condições estandardizadas para cada uma das técnicas. Sempre que as aplicações foram efectuadas numa única sessão, foi providenciado um intervalo entre as duas provas administradas. Foi posto um particular cuidado na pontualidade, no acolhimento e no estabelecimento da relação com os participantes, do que terá, talvez, resultado a excelente atitude face à situação de avaliação da generalidade dos participantes e o interesse que com frequência manifestaram, no final, sobre o tema e a natureza da investigação em curso.

No CAPÍTULO 5 encontra-se uma descrição detalhada dos procedimentos utilizados em cada um dos estudos efectuados.

[31] Agradeço ao Prof. Gerardo Prieto, da Universidade de Salamanca, a sugestão e o precioso apoio à realização deste estudo.

4.2.4. Tratamento dos Resultados: Técnicas Estatísticas

1) **FUNDAMENTAÇÃO DA ESCOLHA E ARTICULAÇÃO DAS TÉCNICAS**

Com a identificação das técnicas de tratamento de resultados procura-se completar a caracterização dos métodos adoptados na presente investigação. Naturalmente, as técnicas foram escolhidas em função das duas grandes etapas do plano experimental, e em particular em função dos propósitos de cada um dos estudos específicos efectuados.

Na primeira etapa – de adaptação e estudo metrológico das técnicas diferenciais – foram utilizados os procedimentos de tratamento de dados consensualmente aceites para este tipo de estudos e enquadrados pela chamada Teoria Clássica dos Testes, por alguns também designada Teoria do Resultado-Verdadeiro (Aiken & Groth-Marnat, 2006; Anastasi & Urbina, 1997; Cronbach, 1970; Guilford, 1954; Guilford & Fruchter, 1978; Kerlinger, 1973; Kline, 1993; Nunnally, 1978; Nunnally & Bernstein, 1994). Procurou-se, contudo, ir além da teoria clássica e integrar as vantagens oferecidas por metodologias mais recentes, muito designadamente pelas que se enquadram na abordagem da Teoria da Resposta ao Item (TRI), também conhecida como Modelos de Traço Latente (Moreira, 2004), em particular, as que se reportam ao "Modelo logístico de um parâmetro de Rasch"[32] (aplicado às medidas compostas por itens dicotómicos) ou ao "Modelo Politómico" ou "Modelo de Rasch de Crédito Parcial", uma extensão do Modelo de Rasch para utilização com *rating scales* (Bond & Fox, 2007, p.123) (aplicado às medidas compostas por itens politómicos, que admitem gradação de cotação). A opção pela inclusão deste tipo de técnica na presente investigação baseou-se nalgumas das vantagens que oferece sobre a teoria clássica dos testes (Embtretson & Reise, 2000; Prieto, Arias-Barahona & Núñez, 2005; Prieto & Delgado, 2003; Prieto & Velasco, 2002; Waugh & Chapman, 2005): a *medição conjunta* (expressão numa mesma escala de medida, a escala *logit*, dos parâmetros dos indivíduos e dos itens); a *objectividade específica* (decorrente da aplicação do princípio fundamental da "comparação invariante"[33] – a diferença entre duas pessoas não depende dos itens com que foram avaliadas, e a diferença entre dois itens não depende das pessoas em que foram avaliados); a utilização de uma *escala de intervalos* (ao longo de toda a escala, uma mesma diferença entre o nível

[32] O modelo logístico de um parâmetro parte do pressuposto de que um único parâmetro (o nível de dificuldade) é suficiente para caracterizar cada item. É de longe o modelo mais utilizado, embora assente numa simplificação ao negligenciar outros parâmetros (como a capacidade discriminativa) que podem caracterizar os itens e estar na origem de deficiente ajustamento ao modelo de um parâmetro (Moreira, 2004).

[33] A "comparação invariante" constitui uma exigência fundamental da medida física: simplificando, a comparação entre dois estímulos deve ser independente dos indivíduos específicos que procedem à comparação e a comparação entre dois indivíduos deve ser independente dos estímulos específicos em que se procede à comparação [Rasch, G. (1961). On general laws and the meaning of measurement in Psychology. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, IV (1980). Berkeley: University of Chicago Press.]. Este princípio é atingido no modelo de Rasch por haver separação algébrica da determinação dos parâmetros das pessoas e dos itens, isto é, o processo estatístico de determinação dos parâmetros dos itens dispensa a consideração dos parâmetros das pessoas e, reciprocamente, o processo estatístico de determinação dos parâmetros das pessoas dispensa a consideração dos parâmetros dos itens.

de competência de um indivíduo e o parâmetro de dificuldade de um item corresponde a uma mesma probabilidade de acerto, ou de resolução com sucesso); e a *especificidade do Erro Padrão da Medida* (EPM) (reconhecimento de que a precisão da medida não é homogénea ao longo de todos os níveis da medida, dos itens fáceis aos difíceis, e possibilidade de quantificação da informação que a medida fornece em cada ponto da dimensão).

Estas vantagens do Modelo de Rasch só se aplicam, contudo, se as medidas se ajustarem ao modelo: o estudo das medidas com esta metodologia consiste, assim, em 1) verificar qual o grau de ajustamento das medidas obtidas numa amostra ao modelo; 2) no caso desse ajustamento não se confirmar, procurar averiguar os motivos do desajustamento e, se possível, corrigi-los; e 3) no caso desse ajustamento se confirmar em grau aceitável, recolher uma diversidade de informações sobre a medida (parâmetros dos itens, isto é, nível de dificuldade independente das características das amostras, função de informação, índices de precisão, dimensionalidade, etc.). Entre os mais frequentes motivos de desajustamento, em testes de resposta múltipla, contam-se a multidimensionalidade, a falta de clareza no enunciado ou nas opções de resposta, a ocorrência de respostas ao acaso, a falta de motivação ou cooperação por parte dos participantes, os erros de anotação da resposta ou a cópia das respostas correctas (Prieto & Delgado, 2003); nos testes de resposta aberta, acrescenta-se o grau de exigência ou a oscilação nos critérios aplicados por juízes (Bond & Fox, 2007).

O modelo logístico de um parâmetro adoptado nesta investigação, o mais clássico da família dos modelos de Rasch, foi proposto em 1960[34] e fundamenta-se em dois pressupostos: primeiro, o atributo que se pretende medir pode representar-se numa única dimensão em que se situam conjuntamente as pessoas e os itens; segundo, o nível de uma pessoa no atributo e a dificuldade de um item determinam a probabilidade de que a resposta ao item seja correcta. A probabilidade de resposta correcta é assim modelada como função logística[35] da diferença entre os parâmetros da pessoa e do item[36], e estes parâmetros expressam-se numa mesma escala intervalar, uma escala logarítmica designada *logit* cujo ponto 0 corresponde, regra geral, ao nível de dificuldade médio dos itens e que, embora seja infinita em ambos os extremos, abrange a maioria dos valores entre -5 e +5.

[34] Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research, citado, por exemplo, em Prieto & Delgado, 2003.

[35] "Logística" no sentido de "lógica algorítmica".

[36] A equação básica do modelo de Rasch é a seguinte: $\boxed{ln\ (P_{si} / 1 - P_{si}) = (\Theta_s - \beta_i)}$ (o quociente entre a probabilidade de resposta correcta, do sujeito $_s$ ao item $_i$, e a probabilidade de resposta incorrecta, do mesmo sujeito ao mesmo item, é função da diferença entre o nível do sujeito no atributo Θ_s e o nível de dificuldade do item β_i. Quando um sujeito responde a um item equivalente ao seu nível de competência, tem a mesma probabilidade de resposta correcta e incorrecta (.50/.50); o *ln* (ou logaritmo natural de) (P_{si} / 1 – P_{si}) reflecte, neste caso (*ln* 1=0), que a dificuldade do item é igual ao nível de competência do sujeito e, por consequência, $\Theta_s - \beta_i = 0$. Se a competência do sujeito no atributo é superior ao nível de dificuldade do item, $\Theta_s - \beta_i > 0$ (probabilidade de resposta correcta superior à probabilidade de resposta incorrecta); se, pelo contrário, a competência do sujeito no atributo é inferior ao nível de dificuldade do item, $\Theta_s - \beta_i < 0$ (probabilidade de resposta correcta inferior à probabilidade de resposta incorrecta) (Prieto & Delgado, 2003).

O modelo de Rasch representa a estrutura que um conjunto de dados deve ter para que a medida de um determinado atributo seja viabilizada por determinado instrumento; consiste num ideal que não é nunca atingido de maneira absoluta, mas que estabelece um formato *standard* contra o qual se torna possível estimar o grau de ajustamento dos dados e a suas propriedades enquanto medida linear de uma dimensão psicológica.

Estas noções de "ajustamento ao modelo" (ou *"fit"*) e de "dimensionalidade" suscitam por vezes alguma confusão desta técnica com os métodos de modelização estrutural. De facto, também com os métodos de "modelização estrutural linear" (*Structural Equation Modeling, SEM*) (Brown, 2006; Kline, 2005), por vezes genericamente designados de métodos estruturais, se procura averiguar o grau de ajustamento de um conjunto de modelos alternativos, em geral derivados da teoria psicológica, à estrutura interna dos dados (que se consubstancia nas matrizes de covarâncias ou de intercorrelações das variáveis medidas). Um modelo estrutural consiste numa formalização matemática que traduz determinadas hipóteses logicamente deduzidas da teoria ou da investigação prévia, relativas aos elementos essenciais de um fenómeno ou às leis que o regem, a qual é confrontada com os resultados empíricos obtidos numa amostra. Uma vez definido o conjunto de variáveis pertinentes para o estudo do fenómeno em causa, o investigador formula hipóteses sobre os efeitos de umas variáveis sobre outras, ou sobre as suas relações recíprocas: alguns efeitos podem ser fixados *a priori*, em função das hipóteses (por exemplo, supondo-se nulos), outros podem estabelecer-se como parâmetros a estimar (muitas vezes recorre-se à representação gráfica para auxiliar a explanação destas relações ou efeitos); uma vez estimados os parâmetros "livres", os que não foram definidos *a priori*, com auxílio de princípios clássicos de estimação (como o dos mínimos quadrados ou o da máxima verosimelhança), verifica-se a adequação ou o ajustamento (o "*fit*") do modelo, comparando a estrutura de dados reconstruída a partir do modelo com a estrutura de dados observada (Bacher, 1987, 1988; Brown, 2006; Kline, 2005). O modelo pode então mostrar-se compatível com os dados, o que numa perspectiva dialéctica de validação (Messick, 1975, 1980) significa que quer a concepção teórica transcrita no modelo sai robustecida pela observação empírica, quer a medição empírica do construto sai legitimada, porque fundamentada pela modelização teórica. Mas o modelo pode mostrar-se pouco ajustado aos dados, e sugerir então ou a comparação com modelos alternativos, que possam alcançar um maior grau de ajustamento, ou a modificação de alguns parâmetros do modelo, de modo a melhorar o seu ajustamento aos dados. Esta última possibilidade apresenta o enorme interesse de conduzir ao aperfeiçoamento do modelo inicial, agora alicerçado na observação empírica, mas torna imprescindível a verificação ulterior do ajustamento do modelo modificado a outras estruturas de dados, emergentes de novos estudos empíricos (Bacher, 1987).

Do problema e das hipóteses atrás formulados decorre, desde logo, a escolha de uma metodologia desta natureza para o tratamento dos dados do Estudo Principal, na presente investigação: a aplicação de técnicas de análise factorial confirmatória, um caso particular das técnicas de modelização estrutural linear (Brown, 2006). Mas torna-se necessário estabelecer se este tipo de método se opõe ou se sobrepõe à análise do modelo de Rasch. Uma diferença fundamental separa desde logo os dois métodos: enquanto na modelização estrutural se procura identificar o modelo que melhor descreve a estrutura dos dados, o que por vezes é alcançado por meio da modificação dos parâmetros do modelo para melhorar o ajustamento, na análise de Rasch o objectivo consiste em obter dados que se conformem ao modelo, isto é, o modelo prescreve um conjunto de restrições que devem ser contempladas para possibilitar a medida de um qualquer construto, num sentido equivalente ao da medição nas ciências físicas. Quando o ajustamento não se confirma, não há que modificar os parâmetros para melhorar o ajustamento do modelo à medida, há que modificar a medida para melhorar o ajustamento da medida ao modelo. Perante um fraco ajustamento, não é o modelo de Rasch que é alterado para melhor se ajustar aos dados; é o método de medida que deve ser revisto e modificado para que satisfaça as exigências do modelo de Rasch, sem o que a sua utilização para a medição do atributo psicológico em causa não sairá legitimada, e as vantagens que o modelo oferece sobre a teoria clássica dos testes não serão aplicáveis. Correndo, embora, o risco de alguma sobresimplificação, pode-se talvez afirmar que a modelização estrutural lida com o estabelecimento do significado da medida (do seu valor conceptual, da sua interpretação, das suas potencialidades de aplicação em avaliação psicológica), ou seja, lida com a validação no quadro da teoria psicológica num determinado domínio, e legitima a *interpretação* e *utilização* da medida. Por seu turno, a "modelização de Rasch" lida com o estabelecimento da qualidade da medida (unidimensionalidade, linearidade, invariância das diferenças), ou seja, lida com a natureza e justificação de determinado método de medição, no quadro de um modelo teórico da medida, e legitima o *método* de medida. A primeira estabelece as potencialidades de generalização da medida a um universo conceptual (validade); a segunda, as potencialidades de generalização da medida a um universo comportamental (generalizabilidade[37]). Nesta óptica, os conceitos de "validade" e de "precisão" abandonam

[37] A teoria da generalizabilidade (TG), introduzida por Cronbach e colaboradores em 1972, pretendeu ultrapassar a concepção estreita de precisão da Teoria Clássica dos Testes (TTC), que se centra na relação entre os resultados observados e os resultados verdadeiros, e questionar antes o grau em que o desempenho num teste permite a generalização para o comportamento da pessoa num universo definido de situações (ou o resultado que obtém permite a generalização para o universo de resultados que obteria, sob todas as condições de observação aceitáveis). Enquanto na TTC a variância de erro é tratada como um todo, na TG ela é decomposta, com recurso à técnica de ANOVA, sendo estimada a magnitude das fontes de erro potencialmente relevantes para o erro de medida, o que permite ao investigador controlar de forma racional o seu efeito em função dos propósitos da medição. Neste sentido, o estudo da generalizabilidade constitui uma extensão do estudo da precisão na TTC (Shavelson, Webb & Rowley, 1989). Mas este estudo da generalizabilidade pode com vantagem basear-se no modelo de Rasch, uma vez que este define teoricamente

decisivamente o tradicional estatuto de distintas "propriedades" ou "características metrológicas" das medidas, para assumirem uma natureza dinâmica, articulada e relacional (bem representativa de uma metametodologia relacional), posto que correspondem a duas perspectivas complementares, ou dois processos inseparáveis, de questionamento da legitimidade de uma medida psicológica.

No número de Janeiro de 1996 da revista editada por Schumacker, *Structural Equation Modeling*, especialmente dedicado à relação entre a análise factorial e o modelo de Rasch, algumas ideias importantes sobressaem: primeiro, de acordo com Wright (1996), quer o não ajustamento ao modelo de Rasch devido a problemas de dimensionalidade, quer as extremidades de variáveis unidimensionais, aparecem na forma de factores menores na análise em componentes principais; segundo, de acordo com Chang (1996), embora ambos os métodos produzam resultados similares, o modelo de Rasch fornece informações de mais fácil interpretação, mais estáveis e mais ricas, posto que esclarece não sobre a proximidade dos dados a uma variável latente, mas sobretudo sobre a localização das pessoas nessa variável, o que poderá facilitar o processo de desenvolvimento teórico; terceiro, de acordo com Smith (1996), quando os dados são dominados por factores altamente correlacionados, ou quando um único factor explica elevada proporção da variância total, o modelo de Rasch mostra-se matematicamente mais apropriado ao estudo da dimensionalidade do que os métodos de análise factorial. Enfim, a conclusão genérica destes e de outros artigos publicados neste número parece favorecer a análise de Rasch, por comparação com a aplicação dos mais clássicos métodos factoriais, sendo sucessivamente demonstrado que os suplanta. O mesmo acontece em alguns artigos publicados numa outra revista que tem dado especial atenção à articulação entre as técnicas de análise factorial e de análise de Rasch, o *Journal of Applied Measurement*, na qual se esboça clara tendência favorável ao modelo de Rasch, quer porque permite detectar falhas nas medidas que escapam por completo à análise factorial (Waugh & Chapman, 2005), quer porque esclarece, e de maneira objectiva ao ultrapassar a arbitrariedade das opções do investigador nos métodos factoriais, as inconsistências ou falta de replicabilidade dos resultados de diferentes estudos factoriais (Kyngdon, 2004). A conclusão do artigo de Smith supracitado sugere, contudo, uma articulação metodológica de maior alcance: a utilização de metodologia factorial exploratória pode servir como ponto de partida para extrair significado das relações internas das medidas, podendo o modelo de Rasch aplicar-se de seguida ao estudo mais específico de determinados conjuntos de itens, que são então justificadamente tratados em conjunto, e examinados quanto ao seu ajustamento global ao modelo e quanto à unidimensionalidade. Este método fornece então a possibilidade adicional de identificar os parâmetros das pessoas e dos próprios itens numa escala intervalar única (Smith, 1996).

(sem necessidade de recurso à recolha de dados empíricos) o nível de precisão máximo que pode ser alcançado como função das características de determinado plano experimental e das inerentes fontes de erro (Linacre, 1993).

A esta articulação metodológica sugere-se que se acrescente que os dados provenientes da análise de Rasch podem revestir-se, por sua vez, de importante significado teórico, ao contribuir para o delinear de modelos conceptuais cujo ajustamento aos dados poderá ser então estudado por meio dos métodos de equações estruturais lineares ou, em particular, pelo método de análise factorial confirmatória. Deste ponto de vista, as técnicas factoriais e de Rasch não se opõem, uma não supera a outra, antes se complementam no estudo metrológico das medidas e em particular no quadro da validação intra-conceito (Dickes, Tournois, Flieller & Kop, 1994). Aos clássicos conceitos de validade e de precisão como propriedades distintas dos testes ou, mais tarde, das medidas, sucede o debate métodos factoriais *versus* modelo de Rasch, tratados como métodos alternativos, posições radicadas numa metametodologia fragmentada; uma metametodologia relacional fundamenta, por seu turno, o tratamento articulados dos dados com recurso a ambos os métodos, entendendo-os como perspectivas complementares que se definem e alimentam reciprocamente. É esta a óptica em que se procuram articular estes dois tipos de técnicas estatísticas no presente trabalho.

2) **IDENTIFICAÇÃO DAS TÉCNICAS DE ANÁLISE DE DADOS**

Na TABELA 4.4 são listadas, por fim, as técnicas estatísticas aplicadas no decurso da investigação. Algumas técnicas fundamentais foram aplicadas nas diversas etapas e no tratamento de variáveis situadas em diferentes níveis (itens, totais, resultados compósitos). Outras técnicas foram aplicadas apenas numa ou noutra etapa do estudo, ou em diferentes estudos e diferentes etapas, com finalidades distintas. As análises estatísticas foram efectuadas com recurso aos seguintes programas informáticos: SPSS 14.0 (SPSS, 2005), AMOS 6.0 (Arbuckle, 2005) (análise factorial confirmatória) e WINSTEPS 3.60 (Linacre,2006) (Modelo Dicotómico de Rasch e Modelo Politómico de Crédito Parcial).

De acordo com o que acima ficou enunciado, a análise de itens foi efectuada no quadro da Teoria Clássica dos Testes (Abordagem Clássica) e também no quadro da Teoria da Resposta ao Item (Abordagem de Traço Latente) – Modelo logístico de um parâmetro de Rasch – modelos dicotómico e politómico; ainda que possa reconhecer-se que a segunda técnica fornece uma leitura dos dados mais rica, e que em larga medida supera a que resulta da aplicação das técnicas tradicionais, mostrou-se necessário tratar os dados também com recurso às técnicas mais clássicas para poder estabelecer algum grau de comparação com dados obtidos noutros estudos. Na aplicação do modelo de Rasch, importa precisar que se utilizou em geral o mais fundamental, que lida com variáveis dicotómicas (1 significa resposta certa e 0 significa resposta errada). Trata-se do modelo que foi atrás descrito e que se designa de "modelo dicotómico". Contudo, para analisar os itens que não admitem apenas respostas certas e erradas mas possibilitam gradação da pontuação, em função da qualidade das (respostas por outras palavras, em que é atribuído crédito parcial por respostas não completamente correctas),

TABELA 4.4
LISTA DAS TÉCNICAS DE ANÁLISE DE DADOS APLICADAS

1ª ETAPA: **Estudos prévios e Ensaios Experimentais**		**2ª ETAPA:** **Estudo Principal**
ANÁLISE DE ITENS	ANÁLISE DE RESULTADOS GLOBAIS (TOTAIS E COMPÓSITOS)	ANÁLISE DA ESTRUTURA DAS MEDIDAS (ITENS E TOTAIS)
Técnicas gerais: -- Estatísticas descritivas (para caracterização das variáveis demográficas e das variáveis experimentais – itens, escalas e resultados compósitos); -- Testes de normalidade das distribuições (Kolmogorov-Smirnov); -- Testes da homogeneidade das variâncias (Levene).		
Teoria Clássica dos Testes: -- estudo da dificuldade dos itens (proporções de acertos); -- estudo do poder discriminativo dos itens (correlações item-escala corrigidas, equivalentes a correlações bisseriais por pontos); -- estudo do contributo de cada item para a consistência interna (Alfa-de-Cronbach, com itens omitidos); -- estudo das taxas de sucesso (nas variáveis politómicas, para ultrapassar perda de informação decorrente da dicotomização); -- estudo da sensibilidade diferencial dos itens à variável sexo – testes não paramétricos (χ^2, Willcoxon Mann-Whitney ou Kruskal-Wallis – amostras independentes); -- estudo da correlação entre as ordenações dos itens original e na amostra portuguesa (coeficiente de correlação ordinal *Rho* de Spearman); -- estudo dos tempos de execução (estatísticas descritivas).	Teoria Clássica dos Testes: -- estudo da consistência interna (Alfa de Cronbach e bipartição); -- estudos diferenciais (variáveis sexo, idade, escolaridade); -- correlações teste-reteste; -- correlações intra-classe (estudo do acordo inter-avaliadores); -- estudo das intercorrelações; -- análise factorial exploratória (eixos principais, rotações ortogonais e oblíquas); -- análise factorial confirmatória (método: máxima verosimelhança).	Teoria Clássica dos Testes: -- Análise das intercorrelações das medidas dos dois instrumentos em conjunto; -- Análise factorial exploratória do conjunto das variáveis (eixos principais, rotações ortogonais e oblíquas); -- Análise factorial confirmatória do conjunto das variáveis (estudo do ajustamento dos modelos derivados das hipóteses experimentais) (método: máxima verosimelhança).
Teoria da Resposta ao Item: -- estudo do ajustamento ao modelo (dicotómico; politómico de crédito parcial); -- estudo dos parâmetros dos itens; -- estudo dos parâmetros das pessoas; -- estudo das opções de resposta (categorias).	Teoria da Resposta ao Item: -- estudo do ajustamento de agrupamentos de itens em resultados compósitos; -- estudo da unidimensionalidade.	Teoria da Resposta ao Item: -- estudo do ajustamento de agrupamentos de itens baseados na análise factorial exploratória ou confirmatória; -- estudo da unidimensionalidade.

foi necessário recorrer ao modelo alternativo, "modelo de crédito parcial"[38]. O tratamento dos dados de parte dos subtestes da *WAIS-III* e dos testes que constituem o *STAT-R* recorreu em geral ao "modelo dicotómico" e nos subtestes da *WAIS-III* com gradação de pontuação em função da qualidade da resposta foi aplicado o "modelo politómico de crédito parcial" (Linacre, 2006). Nalguns subtestes da *WAIS-III* em que existe gradação de pontuação, contudo, os dados foram dicotomizados e de seguida tratados pelo primeiro método, uma vez que a pontuação é efectuada numa escala descontínua que, a ser tratada no quadro do modelo politómico, resultaria em índices distorcidos (por exemplo, no subteste de Cubos em que na maioria dos itens, oito em catorze, a pontuação possível é de 0, 4, 5, 6 ou 7)[39].

Passando da análise dos itens para o nível da análise dos resultados globais, totais ou compósitos, e das suas relações, sobressaiem como técnicas de análise privilegiadas os métodos de análise factorial exploratória e confirmatória. Na primeira etapa da investigação [estudo metrológico das medidas proporcionadas pela *WAIS-III* e pelo *STAT-R (H)*] estas metodologias foram aplicadas sobretudo no quadro da validação intra-conceito, ou estudo da estrutura interna das medidas proporcionadas por cada instrumento; na segunda etapa, preencheram uma função nuclear: a de testar as hipóteses formuladas a partir do problema central sob estudo. Em qualquer das aplicações destes métodos, na primeira ou na segunda etapa, procurou-se não perder de vista a articulação entre a informação decorrente da aplicação das técnicas de análise factorial e a que provém da aplicação da análise de Rasch.

[38] A equação básica deste modelo corresponde a uma adaptação da equação fundamental de Rasch (cf. nota de rodapé 35) que toma em consideração a "calibração" das categorias de resposta (Linacre, 2006).

[39] Para simplificação de linguagem, ao longo do texto os métodos que envolvem os dois modelos – dicotómico e politómico - serão designados genericamente como "análise de Rasch" ou "modelo de Rasch". E a perspectiva metodológica em que se enquadram será designada de "abordagem de traço latente", em complemento da "abordagem clássica", que remete para a chamada "Teoria Clássica dos Testes".